Molecular evolution of the inflated calyx syndrome (ICS) in Withania (Solanaceae)

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Muhammad Ramzan Khan

aus Islamabad/Pakistan

Köln, 2009

Diese Arbeit wurde am

Max-Planck-Institut für Züchtungsforschung in Köln

in der Abteilung Molekulare Pflanzengenetik angefertigt

Berichterstatter:

Prof. Dr. Heinz Saedler

Prof. Dr. Wolfgang Werr

Tag der mündlichen Prüfung**: 21-04-2009**

A necklace made of Withania´s berries

W. somnifera has a long history as a medicinal plant. In the Ayurvedic tradition it is known as ashwaganda (Horse power) or Ayurvedic ginseng and tonic and is used for a wide variety of ailments. Its leaves and berries have been found with Egyptian mummies. (Source: Botanical Museum Berlin)

Table of contents

1 Introduction

1.1 The molecular basis of morphological innovations

Understanding the complexity and diversity of life is still one of the greatest challenges of biology (Theissen et al., 2000; Vergara-Silva, 2003; Carroll, 2005; Theissen, 2006). The appearance of evolutionary novelties, for example new plant parts such as leaves, flowers and seeds are of particular relevance. One of the great evolutionary innovations, the true flower, occasionally, manifests variability in the architecture such as alterations in symmetry, fused sepals and petals and leaf like sepals just to name but a few. These innovations provide organisms with prospects for the exploitation of new ecological niches, which might lead to better adaptive radiation (reviewed in: Mondragon-Palomono and Theissen, 2008). Therefore, delineating the mechanisms underlying morphological novelties are imperative to dissect the biological diversity. Usually, the novel morphological changes in evolution arise by alterations in the developmental processes, which in turn are caused by changes in developmental control genes. However, the relationship between the evolution of genes and morphological structures remain poorly understood. Many of the developmental control genes are members of multigene families and encode transcription factors that play critical roles at key nodes of gene networks, regulating cellular processes. Changes in the number, expression, and interaction of developmental control genes have very likely contributed to the evolution of plant form (Theissen et al., 2000).

1.2 Gene duplication and diversification

Duplicated genes and their diversification are thought to provide the raw material for the evolution of morphological novelties (Force et al., 1999; Carroll et al., 2001; True and Carroll, 2002; He et al., 2004; Harrison et al., 2005). The developmental control genes like other genes can increase in number by gene or genome duplication events, which give rise to paralogues. These paralogues retain the ancestral function shortly after duplication. But during evolution mutations render one of the two copies non-functional because there is no selective pressure to keep both the copies functional. The loss or non-functionalization (pseudogenization) is therefore the ultimate fate of the duplicated genes. However, a large number of duplicated genes escape their extinction, mainly because of sub-functionalization or neo-functionalization (Force et al., 1999; Prince and Pickett, 2002; reviewed in: Mondragon-Palomono and Theissen, 2008). A flow sheet indicating possible events contributing to gene duplication and expected fates of genes after duplication is shown in the

Fig. 1.1. At present in plants, not a single report concerning the positive Darwinian selection and diversification of proteins after duplication in relation to morphological novelties exists, so far.

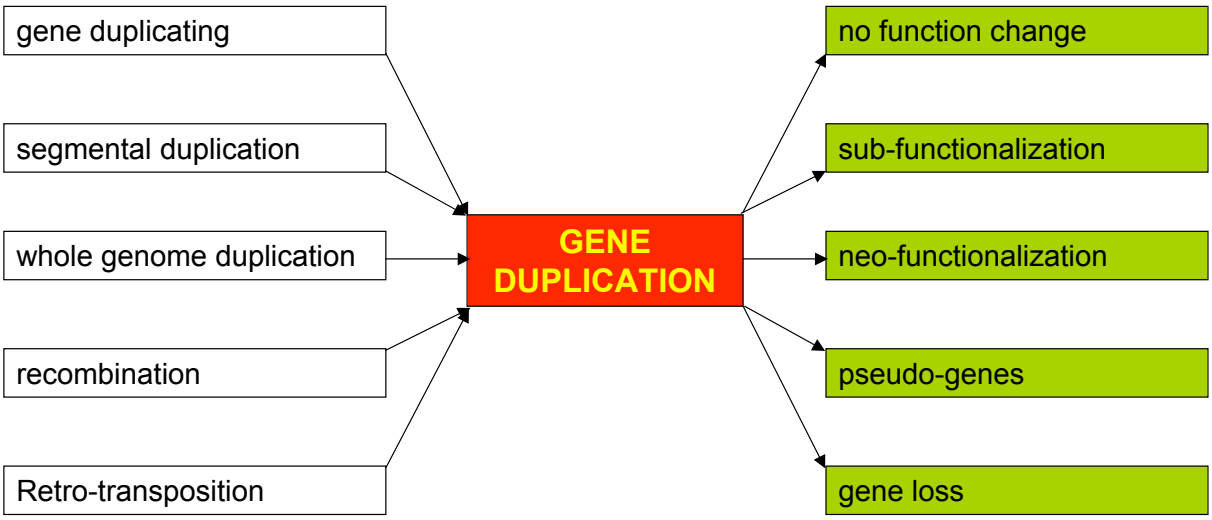


Figure 1.1: Flow sheet showing the contributing factors or types of duplications and the expected fates of genes after duplication

The MADS-box genes family have greatly expanded, particularly in plants, by a range of duplications that have capacitated the genes to diversify in structure and function (reviewed in: Rijpkema et al., 2007).

1.3 The MADS-box transcription factors

A very well known family of developmental control genes is the family of MADS-box genes. Members of the MADS-box gene family (Schwarz-Sommer et al., 1990) encode transcription factors, which play fundamental roles in developmental control and signal transduction in a diverse range of eukaryotic organisms including yeasts, plants, and mammals. MADS-box genes control diverse developmental processes in flowering plants ranging from root to flower and fruit (Shore and Sharrocks, 1995; Theissen et al., 1996; Riechmann and Meyerowitz, 1997; Theissen et al., 2000; Ng and Yanofsky, 2001; Theissen, 2001).

In flower some MADS-box genes act as floral organ identity genes (Schwarz-Sommer et al., 1990; Yanofsky et al., 1990; Theissen et al., 2000). In addition to floral homeotic functions,

MADS-box genes play diverse roles in gene networks that carry out reproductive development in flowering plants (Riechmann and Meyerowitz, 1997; Theissen and Saedler, 1999; Theissen et al., 2000, Ng and Yanofsky, 2001; Theissen, 2001). Some MADS-box genes act as flowering time genes which, depending on internal or environmental factors repress or promote the floral transition (Michaels and Amasino, 1999; Sheldon et al., 1999; Hartmann et al., 2000; Lee et al., 2000; Sheldon et al., 2000). Flowering time genes exert their function by influencing meristem identity genes, which control the transition from inflorescence to floral meristems. Within floral meristems, the boundaries of floral organ identity gene functions are established by cadastral genes, thus defining the different floral whorls (reviewed in: Becker and Theissen, 2003).

MADS-box genes are also involved in seed and fruit development. The protein products of some genes are required for the proper development of the fruit dehiscence zone of Arabidopsis, while another is required for the normal pattern of cell division, expansion, and differentiation during morphogenesis of the silique (Gu et al., 1998; Liljegren et al., 2000; de Folter et al., 2006). Moreover, expression of a number of MADS-box genes outside flowers and fruits as well as an increasing number of mutant and transgenic flowering plants suggests that members of this gene family play regulatory roles also during vegetative development, such as embryo, root, or leaf development (Ma et al., 1991; Rounsley et al., 1995; Huang et al., 1995; Alvarez-Buylla et al., 2000a; Theissen et al., 2000; He and Saedler, 2005; Danilveskaya et al., 2008). The MADS-box genes are also found in lower plants such as in gymnosperms, ferns, and mosses, which do not form flowers or fruits. This further demonstrates that these genes are recruited to functions other than flower or fruit development (Tandre et al., 1995; Münster et al., 1997; Winter et al., 1999; Krogan and Ashton, 2000; Münster et al., 2002b; Henschel et al., 2002; Riese et al., 2004; Kwantes, unpublished results). The members of *STMADS11* gene clade play a role in vegetative development. For example *SVP* is expressed in leaves and shoot apical meristems while *AGL24* is strongly expressed in apical meristem and leaf primorida. The expression of potato *STMADS16* is restricted to leaves while its ortholog *MPF2* in Physalis is expressed in flower as well (He and Saedler, 2005). In addition to leaf and flower development MADS-box genes are also involved in fertility pathways (He and Saedler, 2005).

The MADS-box represents an acronym of the four founding members *MINI CHROMOSOME MAINTENANCE 1* (*MCM1*) in yeast, *AGAMOUS* (*AG*) in Arabidopsis, *DEFICIENS* (*DEF*) in Antirrhinum and Serum Response Factor (*SRF*) in humans (Schwarz-Sommer et al., 1990). All the MADS-box genes share a highly conserved 180 bp long nucleotide sequence, the

MADS-box, encoding the DNA-binding domain of MADS-domain transcription factors (Shore and Sharrocks, 1995). Outside the DNA-binding domain, the proteins are rather diverse (Alvarez-Buylla et al., 2000b).

MADS-domain proteins regulate expression of target genes by binding to CArG-box motifs in their promoter regions (Egea-Cortines et al., 1999; Honma and Goto, 2001; Theissen and Saedler, 2001; reviewed in: de Folter et al., 2006; Verelst et al., 2007). Most MADS-domain proteins prefer the so-called Serum Response Element (SRE-) type CArG-box with a consensus of CC(A/T)6GG sequence (Hayes et al., 1988; Riechmann et al., 1996).

Based on their sequence similarity within the MADS-domain, eukaryotic non-plant MADS-box genes have been subdivided into two groups, Serum Response Factor (SRF) type and Myocyte Enhancer Factor-2 (MEF2) type with various regulatory functions (Theissen et al., 1996; reviewed in: Theissen et al., 2000). By contrast, the majority of the described plant MADS-domain proteins belong to the MIKC-type that is characterized by a conserved structural organization. Downstream of the highly conserved MADS-domain (M-) they feature an intervening (I-) domain, responsible for specification of dimerization (Fan et al., 1997), followed by a conserved keratin-like (K-) domain that promotes protein dimerization (Riechmann and Meyerowitz , 1997; Kaufmann et al., 2005), and a C-terminal domain (C-) that contributes to the transcriptional activation and formation of multimeric complexes (Egea-Cortines et al., 1999; Honma and Goto , 2001; Kaufmann et al., 2005). Some MIKC proteins possess an additional N-terminal extension (Theissen et al., 1996).

The remaining plant MADS-domain genes are of the M-type and lack the plant-specific K-domain. Their functions are largely unknown. Alvarez-Buylla et al., (2000b) divided SRF, MEF2, and MIKC-type MADS-box genes into two groups, type I and II, according to a phylogenetic analysis of the MADS-box. According to this, type I includes animal and fungal SRF-type genes as well as a group of largely uncharacterized genes, among them all plant M-genes. Type II consists of MEF2-type genes and the MIKC-type genes that are unique to plants. Parenicova et al., (2003) suggested a further subdivision of all plant M-type MADS-box genes into four groups termed Mα-Mδ, while Mδ-type genes consist exclusively of MIKC*-type genes (Kofuji et al., 2003). Phylogenetic analyses revealed that, due to a remarkable similarity between the MIKC^c ("classic") and the MIKC*-type genes, the latter fit within the type II group instead of type I.

MIKC*-type genes differ from the MIKC^c-type in the length of the I-domain and composition of the K-domain (Henschel et al., 2002). Their functions are largely unknown (Verelst et al., 2007). The only MIKC MADS-domain gene that was described for each of

three Charophycean green algae, representatives of the closest living relatives of land plants (Karol et al., 2001), was of the classic type (Tanabe et al., 2005).

Theissen et al., (1996) further sub-divided the MIKC^c-type genes into several well-defined gene clades, termed 'gene subfamilies' to have a rigorous and unambiguous system for subfamily nomenclature. These are always named after the first clade member that has been identified. Members of the early identified and well-characterized subfamilies, such as the *AG*, *DEF*, *GLO*, and *SQUA-like* genes, typically share similar expression patterns and highly related functions (Doyle, 1994; Purugganan et al., 1995; Theissen and Saedler, 1995; Theissen et al., 1996; Theissen et al., 2000). There are 12 gene families which include: AG-like genes; AGL2-like genes; AGL6-like genes; AGL12-like genes; AGL15-like genes; AGL17-like genes; DEF+GLO-like genes; FLC-like genes; GGM13-like genes; SQUA-like genes; STMADS11-like genes and TM3-like genes.

The establishment of the different gene clades as mentioned previously by gene duplication, diversification, and fixation is important for the establishment of the floral homeotic functions, and thus flowers (Theissen et al., 1996; Theissen et al., 2000). From these gene families it is clear that MADS-box transcription factors are involved in the diverse developmental processes such as vegetative and floral development.

The best-characterized MIKC^c-type MADS-box genes are those that function as floral organ identity genes in angiosperms (Schwarz-Sommer et al., 1990; Weigel and Meyerowitz, 1994).

1.4 MADS-box genes and flower development

Eudicots have relatively standardized flowers that typically, are composed of four distinct organ types that arise in concentric rings, referred as whorls of the organs. The outer two whorls contain sterile organs that make up the perianth. Sepals arise in the first whorl and form the outer covering of the developing bud. The sepals protect the developing flower buds against adverse biotic and abiotic effects, such as insect damage and temperature changes. Petals arise in the second whorl and are often large, colourful and conspicuous to attract pollinators. In some flowering plants the first and second whorl organs have the same form and are known as tepals. The inner two whorls of organs are needed for reproduction. Stamens arise in the third whorl and produce pollen, which develop into the male gametophytes. Carpels, which are also known as pistils arise in the central fourth whorl and produce ovules that contain the female gametophytes, the embryo sac. Often the carpels are united or fused together to form the gynoecium. Once fertilized, the gynoecium develops into

the fruit harboring the seeds, although in some species the fruits develop from other parts of the flower, such as the receptacle.

Genes encoding MADS-box transcription factors define the identity of these four types of organs (Schwarz-Sommer et al., 1990; Yanofsky et al., 1990; Theissen et al., 2000). These genes encode transcription factors that activate or repress target genes in their distinct expression domains, thereby define the whorls of the angiosperm flower. Floral homeotic mutations, in which one type of organ is replaced by another type of organ, have been described in various species. It is worth mentioning that one mutational step can change the identity of organs in two adjacent whorls.

The genetic analysis of floral homeotic mutants in *Arabidopsis thaliana* and *Antirrhinum majus* in 1991 led to the formulation of ABC model for the specification of floral organ identity (Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994).

1.5 The ABCDE model

According to the ABC model three regulatory gene functions i.e. A, B and C work in a combinatorial fashion to specify the organ identity in each floral whorl. In Arabidopsis the A function is specified by class A genes *AP1* (*APETALA1*) and *AP2* (*APETALA2*) while, A and B class genes *AP3* (*APETALA3*) and *PI* (*PISTILLATA*) specify the petals in the second whorl. The *DEF(DEFICIENS)* and*GLO* (*GLOBOSA*) genes are responsible for B function in Antirrhinum. The B and C class genes (*AGAMOUS*) define stamens (male reproductive organs) in the third whorl, and C class genes specify carpels (female reproductive organs) in the fourth whorl of a typical Arabidopsis flower (Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994; reviewed in De Bodt et al., 2006). The counterparts of *AGAMOUS* in Antirrhinum are *PLE* (*PLENA*) and *FAR* (*FARINELLI*). The ABC model is basically applicable to all the widely studied species such tulip, petunia, primrose, rice and maize though details differ (Angenent et al., 1992; Kanno et al., 2003; Webster and Gilmartin, 2003; Nagasawa et al., 2003). Studies of genetic redundancy have resulted in the identification of E class genes and ultimately led to the extension of ABC model. These E class or *SEPALLATA* genes (*SEP1, SEP2, SEP3, SEP4*) in Arabidopsis are redundantly involved in the specification of all floral whorls as revealed by the triple and quadruple mutants in Arabidopsis (Ditta et al., 2004). The *SEP* orthologs such as *FBP2* and *FBP5* (*FLORAL BINDING PROTEIN 2* and *5*-) are also found in other plants species like in petunia (Vandenbussche et al., 2003b). In addition to ABCE class genes, D class genes (*SEEDSTICK*

and *SHATTERPROOF 1* and *2*), important for ovule development have been identified in Arabidopsis (Theissen et al., 2001).

1.6 The "Floral Quartets " model

Most ABC function genes encode MADS-box transcription factors. These proteins form dimers (Troebner et al., 1992; Davies et al., 1996) and tetramers of different composition to bind to their CArG (CC (A/T)6 GG) motifs (Egea-Cortines et al., 1999; Honma and Goto, 2001). This finding that MADS domain protein can associate in higher order complexes for functioning led to the formulation of a model called "floral quartets" (Theissen and Saedler, 2001). Different "Quartets" are involved in the establishment of floral organ identity (Fig. 1.2). According to this model, the identity of the different floral organs (sepals, petals, stamens and carpels) is determined by four combinations of floral homeotic MADS-box proteins. In each whorl, dimers of floral MADS proteins are proposed to bind to CArG (CC (A/T)6 GG) motifs in the promoter of their target genes. These DNA sites could be in close vicinity or some distance apart along the DNA. According to the model, tetramers are formed through protein-protein interactions between the MADS proteins dimers. This whole complex is bound to two CArG-boxes in the DNA. Figure 1.2 shows the predicted composition of

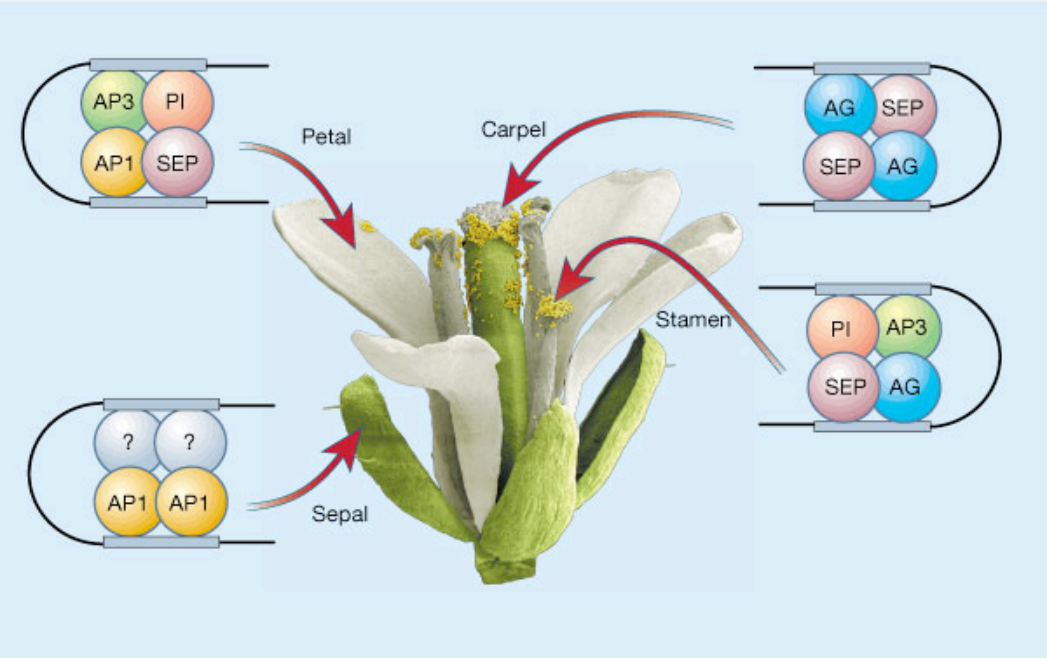


Figure 1.2: **Flower structure and the 'quartet model' of floral organ specification in Arabidopsis.** Question marks denote components whose identity is not known. (Adopted from Theissen and Saedler, 2001)

tetramers in the four whorls. AP1 can combine with AP1 and other two unknown proteins (SEP may be) in whorl 1 to specify sepals. The complex for the specification of petals is AP1-SEP-AP3-PI, while AG-SEP-AP3-PI quartet specify stamens. For carpels the complex AG-AG-SEP-SEP is proposed in the whorl 4. The empirical basis of floral quartet model is still quite weak, because there is no direct evidence that multimeric complexes of floral homeotic proteins really exist and how target gene specificity is achieved by these proteins (Melzer et al., 2006).

MADS-box genes are essential for floral organ identity. Mutations within these genes may cause structural novelties in flower. It is important to describe in elaborate details the potential significance of MADS-box genes in generating floral novel morphologies.

1.7 On the origin of floral morphological novelties

As previously mentioned, the origin of morphological novelties has long been a challenge in evolutionary biology. Various developmental and genetics mechanisms give rise to such structures. It is found that heterotopic expression of pre-existing functions rather than the origin of new functions is responsible for these traits in animals and plants (Wang et al., 1999; Keys et al., 1999; Frary et al., 2000; Davies et al., 1999; Lee et al., 2003; Hileman et al., 2003; Kanno et al., 2003; He and Saedler, 2005). In order to generate the novel floral morphologies, changes in *cis*-acting (Doebley and Lukens, 1998) or alterations in transacting transcriptional regulators are considered to be involved usually.

MADS-box genes specify the floral organ identity. Other transcription factors, like TCPs, are involved in determining the symmetry of the flower and especially, the symmetry of petals in whorl 2 (Luo et al., 1996; Luo et al., 1999). The idea about the floral morphological novelties came from *Antirrhinum majus* where bilateral symmetry is maintained by *Cycloidea* and *Dichotoma*, which are *TCP* transcription factors. Mutation in one or both of the genes leads to semi-radial and perfectly radial flowers (Fig. 1.3: 1, 2a, 2b). Radial flower is also generated by *CYC* inactivation (Fig. 1.3: 3a, 3b).

According to the model different “quartets” are involved in floral organ identity. The synthesis of new “quartets” in a given whorl should lead to a change in the identity of the organs in that whorl (reviewed in: He et al., 2004). For example, if B-function genes, like *DEF* and *GLO*, were expressed ectopically in the first whorl, then petals should replace sepals and hence such “mutants” should reveal two whorls of petals as has been shown for transgenic *Nicotiana tabacum* (Davies et al., 1996) (compare Fig. 1.3: 4a and 4b). Petals instead of sepals already exist in nature. An example is tepals instead of sepals and petals in

tulips (Fig. 1.3: 5). Molecular analysis has revealed that this structure is generated due to ectopic expression of B-function (Theissen et al., 2000; van Tunen et al., 1993; Kanno et al., 2003). Within Rubiaceae both leaf-like and coloured (petal-like) sepals are found (Fig. 1.3: 6, 7, 8). Leaf-like sepals or foliose-sepal-syndrome (FSS) as the structure is often called, have been isolated as mutant in several model species. These include *Zea mays*, *Solanum lycopersicon* and *Arabidopsis thaliana*.

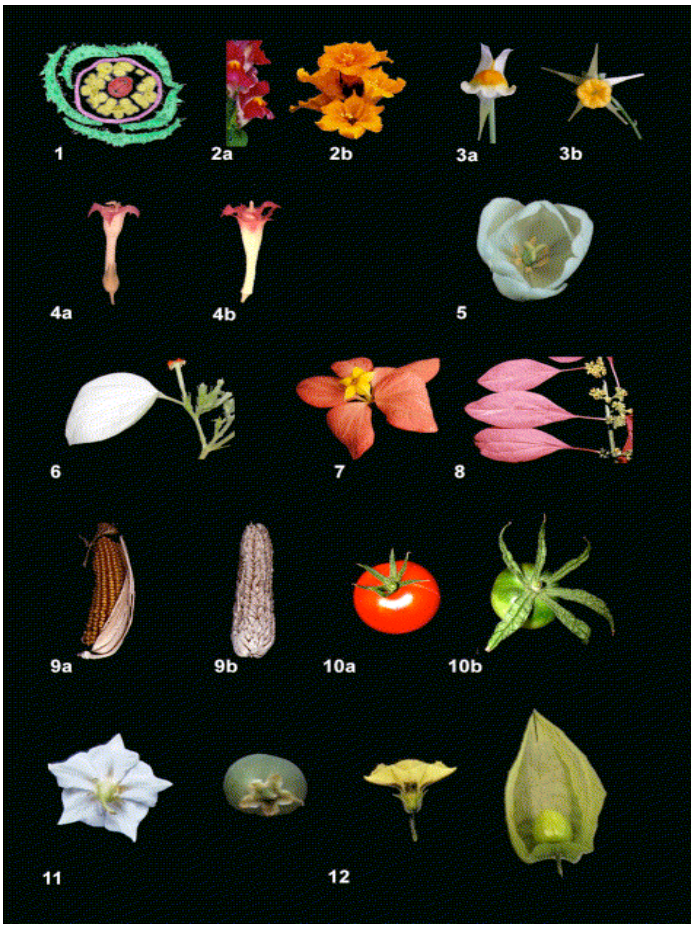


Figure 1. 3: Diversity in floral morphological novelties in different species. 1) Cross-section through a flower of *Antirrhinum. majus*. 2a) Wild type flowers of *A. majus*. 2b) Phenotype of a double mutant cyc/dich featuring radial symmetry. 3a) Zygomorphic flower of *Linaria vulgaris*. 3b) Epigenetic cyc mutation of *L. vulgaris* resulting in an actinomorphic flower. 4a) *Nicotiana tabacum* wild type flower. 4b) Transgenic *N. tabacum* overexpressing the B-function genes from *A. majus*. 5) Flowers of *Tulipa gesneriana.* 6) Flower of *Mussaenda frondosa*. 7) *M. erythrophylla* red leaves replace the 5 sepals. 8) *Warszewiczia coccinea* a red leaf replaces one sepal. 9a) Kernels on a cob of *Zea. mays*. 9b) Kernels covered by glumes in Tunicate mutation. 10a) *Solanum lycopersicum*. 10b) lemads-mc mutant shows leaf-like sepals. 11) Flower and fruit of *S. tuberosum*. 12) Flowers and fruit of *Physalis. floridana*. (Adopted from He et al., 2004)

Molecular analysis of Tunicate mutation in *zea mays* revealed that MADS-box gene *ZMM19*, whose expression in the wild type is confined to leaves and husks, is now expressed ectopically in all floral organs. The ectopically expressed *ZMM19* protein seems to promote growth of the glumes, which then ultimately cover the entire kernel (compare Fig. 1.3: 9a and 9b). Even though glumes are not necessarily homolog to sepals, they might serve as a model for the type of genes involved in FSS of higher eudicots (reviewed in: He et al., 2004). This assumption is corroborated by the observation that *ZMM19* ectopic expression in *A. thaliana* leads to foliose sepal formation. *ZMM19* is a member of the *STMADS11*-clade (Garcia-Maroto et al., 2000; Münster et al., 2002a) of MADS-box genes and genes belonging to this clade may lead to a higher proliferation potential of vegetative tissues (Kim et al., 2002).
Mutation in the *LeMADS-MC* gene affects sepal size in *Solanum lycopersicum* (compare Fig. 1.3: 10a and 10b) (Vrebalov et al., 2002). Unlike *ZMM19* of *Z. mays* described above, *LeMADS -MC* belongs to the SQUA-clade of MADS-box genes (Huijser et al., 1992). These results indicate that two different MADS-box genes can generate FSS, either ectopic expression of a vegetative *STMADS11-like* gene or a knockout mutation of a *SQUA-like* gene (Fig. 1.3: 10b). Knockout mutations of the *AP1* gene of *A. thaliana*, an ortholog of *SQUA*, showed a complex phenotype including large or foliose sepals (Mandel et al., 1992). *AP1* is a repressor of *AGL24*, *SVP*, and *SOC1*, which are *STMADS16-like* genes (Liu et al., 2007). If mutated, the vegetative expression of these genes extends into floral development thereby generating foliose sepals. Ectopic expression of *STMADS16*, *MPF2*, *AGL24*, *SVP* and *ZMM19* genes, all belonging to the *STMADS11*-clade, results in FSS in *A. thaliana* (He and Saedler, 2007) and Nicotiana (Garcia-Maroto, 2000).
The above molecular studies indicate that novel floral morphologies not only exist in nature but also can be generated by genetic manipulations.

1.8 The inflated calyx syndrome (ICS): a floral morphological novelty in Solanaceae

Among angiosperm families, the Solanaceae rank as one of the most important to human beings. Species of the family are used for food, drugs and ornamentals. It is rich in flower and fruit diversity (Knapp, 2002; Knapp et al., 2004). The previous section highlights that flower is an attractive target for the evolutionary studies of novel morphologies. The floral calyx of Solanaceae seems to be a playground for evolution of morphological novelties. A tremendous degree of calyx diversity exist, from small fused sepals as in *Solanum tuberosum* to small sepal rings as in *Tubocapsicum anomalum*, to large plate-like calyces as in *Atropa belladonna*, to leaf-like sepals as in *Calibrachoa parviflora*, to long tubular calyces as in

Brugmansia aurea or to lantern-like calyces as in Physalis or in Withania species. The calyx of the Physalis flower resumes growth post pollination and the resulting ICS or "Chinese lantern" as the structure is often called, encloses the developing berry. An inflated calyx syndrome (ICS) is characteristic for quite a number of species within several genera of the Solanaceae (D'Arcy, 1991; He and Saedler, 2004). Among these Physalis, Withania and Nicandra are the spectacular ones. The function of the inflated calyx is not entirely clear, but in certain species like *Przewalskia tangutica* it seems to reduce the specific weight and thus might facilitate wind dispersal of the fruits (Knapp, 2002).

1.9 Withania

Withania is a shrub or perennial herb, with erect and much branched stem. Leaves are solitary or paired but simple. Fruiting calyx become enlarged, enveloping berry, and closed at the apex. Fruit is a globose shiny berry and seeds are compressed. It is a genus of about 11 species, six (*W. qaraitica, W. riebeckii, W. adunensis, W. reichenbachii, W. sphaerocarpa, W. grisea*) endemic in the Horn of Africa region, three (*W. adpressa, W. aristata, W. frutescens*) in the Canary Islands, North Africa and Spain, one (*W. coagulans*) in southern Asia, and one (*W. somnifera*) species that is widespread in the arid zones of tropics and subtropics of the Old World. *W. somnifera* has a long history as a medicinal plant. In the Ayurvedic tradition it is known as ashwaganda (Horse power) or Ayurvedic ginseng and tonic and is used for a wide variety of ailments. Its leaves and berries have been found with Egyptian mummies. *W. somnifera* is also much more variable than the other species and is known to include polyploid forms.

1.10 Tubocapsicum

The plants of *T. anomalum* are erect or sprawling herbs with entire leaves, few-flowered inflorescence with small campanulate flowers, and completely exposed red juicy berries.

Tubocapsicum, a genus of two Asian species, was formerly included in the Capsicum. The genus Tubocapsicum is endemic to eastern Asia. It embraces one wide spread species, *T. anomalum* and a poorly known second species, *T. obtusum*. The genus prefer humid climate. It is not well known since there has been little published information, and there are few specimens in western herbaria (D'Arcy et al., 2001).

1.11 The molecular basis of ICS

He and Saedler (2005) did the pioneer work on the molecular basis of the ICS formation in Physalis. In an elegant study they demonstrated that a MADS-box gene *MPF2*, an ortholog of *STMADS16* in potato, is involved in the ICS formation in *Physalis floridana*. Recently, Hu and Saedler (2007) constructed a phylogeny of the Solanaceae based on the two chloroplast markers *matK* and *atpB* and the two nuclear markers *MPF1-like* and *MPF2-like*. They demonstrated that Withania and Tubocapsicum are sister to each other (Fig. 1.4).

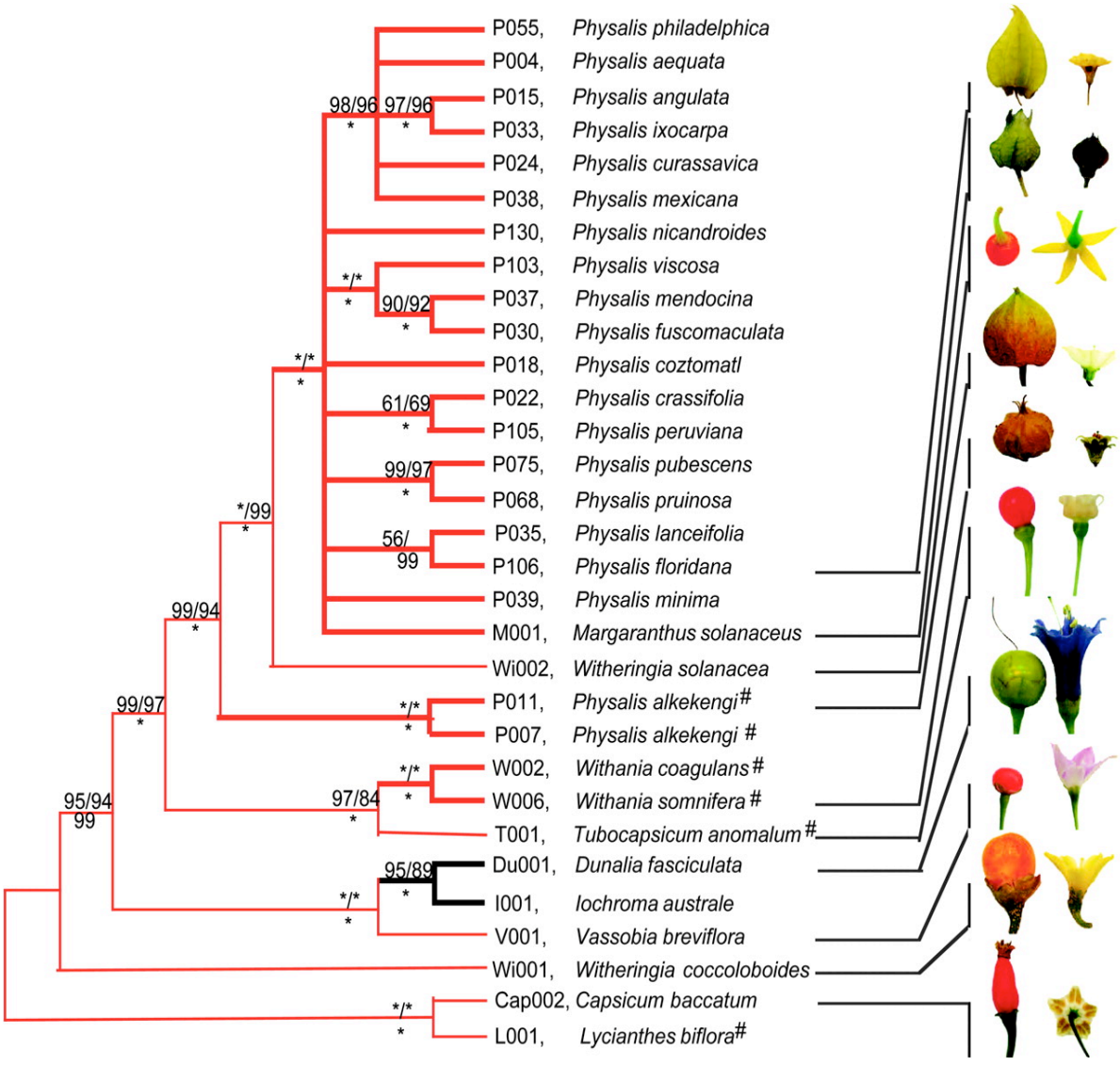


Figure 1.4: **Phylogenetic tree of the Physaleae.** A strict consensus Maximum Parsimony tree with Capsicum and Lycianthes as out-groups based on concatenation of the 2 chloroplast sequences *atpB* and *matK* and cDNA sequences of the nuclear *MPF1-like* and *MPF2-like* genes is shown. The flower and fruits of a few plant species are also shown. (Adopted from Hu and Saedler, 2007)

This finding is very interesting as Withania has a lantern comparable in inflation to Physalis while on the other hand Tubocapsicum features only a rudimentary calyx. How these two genera with such a different morphologies with respect to ICS formation, become closer relative to each other, remains a challenging question.

Previously it is shown that the expression of *MPF2* in the calyx is a pre-requisite for ICS formation. However, this is not the only component involved but the two plant hormones cytokinins and gibberellins are also essential (He et al., 2004; He and Saedler, 2005). These hormones are induced in response to fertilization. Hence, they mimic the fertilization signals. Cytokinin facilitates the transport of MPF2 to the nuclei of calyx cells thus promoting cell division resulting in small cells, which in the presence of gibberellin enlarge and ultimately form ICS (He and Saedler, 2007). However, this is not the only function of MPF2. 35S::MPF2 RNAi knock-down plants in which the level of *MPF2* RNA was reduced to ca. 5% featured small leaves, no ICS and were male sterile (He and Saedler, 2005). This suggested that MPF2 has multiple functions: in leaf and calyx development and in male fertility.

MPF2 is known to interact with other MADS-box proteins to form hetero- or even higher oligomers (He et al., 2007). These might control cell division in either leaves or the calyx depending on their interacting partners. However, MPF2 interacts also with PFMAGO, an ortholog of mago nashi from Drosophila known to control formation of progeny (Boswell et al., 1991; Newmark and Boswell, 1994), a role it might serve in Physalis as well (He et al., 2007).

Recently, floral expression of *MPF2-like* genes in most Physaleae and most Solanaceae tested was shown to be plesiomorphic suggesting that the early Solanaceae might not only have expressed *MPF2-like* genes in floral tissues, but due to this might have featured ICS as well. Later in evolution of the family secondary losses seem to have occurred generating the diversity currently observed (Hu and Saedler, 2007). As suggested before (He and Saedler, 2007), the synthesis of ICS involves many steps including expression of MPF2-like proteins in the calyx, synthesis of cytokinin and gibberellin in the fertilized ovule and their transport to the outer whorl of the calyx. Then MPF2 is modified in that tissue and transported to the nucleus of the cells, its interaction with appropriate partner molecules bring about the cell division, cell elongation and ultimately the ICS. This pathway could be intercepted or modified at many sites. Thus far, only three “evolutionary” mutants (species without ICS) were found blocking the expression of *MPF2-like* genes in calyx tissue: *Solanum tuberosum* (He and Saedler, 2005), *Dunalia fasciculata* and *Iochroma australe* (Hu and Saedler, 2007).

Several other species still expressed their *MPF2-like* gene in calyx tissue but did not feature ICS. Only in the case of *Witheringia coccoloboides* could the defect roughly be placed within the pathway. In all other cases tested, so far the block remained unassigned.

In most diploid Physaleae species *MPF2-like* is a single copy gene (He and Saedler, 2005) and hence due to the importance of its functions, especially in leaf development and in male fertility, the identification of "evolutionary" mutants, i.e. species deficient in one or the other function is highly unlikely. However, polyploid species could provide the alternative material to isolate such "evolutionary" mutants with separated MPF2 functions. At least three sub-tribes for the Physaleae were suggested based on molecular phylogenetic tree reconstructions: Iochrominae, Physalinae and Withaninae. The latter consists of nine genera including Withania and Tubocapsicum (Ohlmstead and Bohs, 2007). While Withania species feature ICS, *Tubocapsicum anomalum* does not. This latter species has only a small inconspicuous ring of calyx (Hu and Saedler, 2007).

Since polyploidy is common within Withania and Tubocapsicum species (Baquar, 1967; Bir and Neelam, 1980; Iqbal and Datta, 2007), hence these plants provide the best suitable materials for studying the separated functions in terms of sub-functionalization and/or neo-functionalization of *MPF2-like* genes within the framework of ICS evolution.

1.12 Aims of the thesis

The main aim of thesis was to investigate the molecular basis of ICS novelty prevailing in Withania. Another major goal of the study was to contribute to understanding the sub-functionalization and/or neo-functionalization of *MPF2-like* genes in the Withaninae and, thereby, provide insight concerning the evolution of ICS. For this purpose, *MPF2-like* genes were isolated, analyzed and their phylogenetic network was determined. The functional divergence was ascertained by over-expressing the *MPF2-like* genes in Arabidopsis as well as finding their interacting partners using the yeast system. An additional investigation of putative regulatory functions of the promoter and first intron *cis*-elements was performed to reveal the regulation of the *MPF2-like* genes expression. Based on the phylogenetic analysis and functional data, sub-functionalization and/or neo-functionalization of the *MPF2-like* genes will be discussed within the framework of ICS formation. Finally, the results will be embedded into an evolutionary context.

2 Materials and Methods

2.1 Sources of supply for commercially available materials

The chemical products used for the preparation of buffers, media and solutions were obtained from Sigma-Aldrich (München), Clontech (Heidelberg), Biozym (Rockland, USA), Invitrogen (Karlsruhe), Merck (Darmstadt), Promega (Heidelberg), Qiagen (Hilden), Roche (Mannheim), Roth (Karlsruhe), Serva (Heidelberg) and Duchefa (Haarlem, the Netherlands).

The enzymes were purchased from KMF Laborchemie Handels GmbH (Lohmar), New England Biolabs (Schwalmbach), Takara Bio Inc. (Otsu, Japan), Roche (Mannheim) and Invitrogen (Karlsruhe). All the enzymatic reactions were performed according to manufacturer's protocols. For normal cloning and sequencing pGEM-T Easy vector from Promega (Mannheim) was used.

The radioisotope [α^{32} P]-dCTP (10 μCi/mol) was obtained from Hartmann-Analytik (Braunschweig).

The commercial kits used for clean-up procedures of nucleic acids, isolation of plasmid DNA and 5'RACE (Rapid amplification of cDNA ends) were obtained from Qiagen (Hilden) and Roche (Karlsruhe).

Molecular weight markers for nucleic acid fragment size determination in gel electrophoresis were obtained from Eurogentec Deutschland GmbH (Köln), New England Biolabs (Frankfurt am Main.) and Fermentas (Baden-Württemberg).

Sigma-Genosys (Steinheim), Invitrogen (Karlsruhe), Operon (Köln), and Metabion International AG (Martinsried) synthesized the oligonucleotide primers used in this work. Some of the primer sequences are listed in the supplement section.

2.2 Media, buffers and solutions

All the media, buffers and solutions were prepared according to Sambrook and Russell (2001) unless mentioned otherwise.

2.3 Scientific softwares and online tools

A number of scientific softwares and online tools were employed to analyse the data, especially the sequences (Table 2.1). The softwares used included: MacVector™ 7.2.3 (Accelrys Inc.) gcg/Wisconsin Package University of Wisconsin); AssemblyLIGN™ (Oxford Molecular Group Plc.); Image Quant (Molecular Dynamics, Krefeld) and ImageJ http://rsb.info.nih.gov/nih-image).

Tool	Application	URL
Rebase	Restriction enzymes	http://rebase.neb.com/rebase/rebase.html
NCBI (BLAST & PUBMED)	Database for sequence information and literature	http://www.ncbi.nlm.nih.gov/
SolGD	Solanaceae Genomics Network	www.sgn.cornell.edu/
NetPlantGene Server	Splice site prediction	http://www.cbs.dtu.dk/services/NetPGene/index.php
PLACE	*cis*-regulatory sequences	http://www.dna.affrc.go.jp/place/
MULAN	Multiple-sequence local alignment	http://mulan.dcode.org
TRANSFAC	*cis*-regulatory sequences	http://www.gene-regulation.com/pub/databases.html
MultiTIF	Alignments and transcription factor binding sites (TFBs) determination	http://rvista.dcode.org/cgi-bin/mTTP.cgi

Table 2.1: Online tools used for *in silico* analyses of nucleic acid and protein sequences

2.4 Bacterial strains and plasmids

E. coli strain DH5α was routinely used for normal cloning and constructs making. Other bacterial strains and plasmids were used according to the nature of experiment (Table 2.2).

Strain	Genotype
E. coli DH5α	SupE44 ΔlacU169 (80 lacZΔM15) hsdR17 recA1 EndA1 gyrA96 thl-1 relA1
E. coli DH10B	F-mcrA D (mrr-hsdRMS-mcrBC) F80diacZDM15 DlacX74 endA1 recA1 D (ara, leu) 7697araD139 galU galK nupG rpst T1R
E. coli DB3.1	F- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(rB-, mB-) ara14 galK2 lacY1 proA2 rpsL20(Smr) xyl5 Δleu mtl1
E. coli TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-
A. tumefaciens LBA4404	pAL4404RS
A. tumefaciens GV3101	pMP90RK
A. tumefaciens GV3101	pMP90RG

Table 2.2: Genotypes of bacterial strains

2.5 Isolation of nucleic acids

Plant genomic DNA was isolated using CTAB (Doyle and Doyle, 1987, 1990) and DNeasy Plant Maxi kit (Qiagen). Plasmid DNA was isolated using the Mini/Midi Plasmid DNA purification kit (Qiagen) according to manufacturer's specifications. PCR clean up and agarose gel extraction was performed with NucleoSpin®ExtractII (Macherey-Nagel).

The quality and quantity of DNA was judged by the comparison of band intensity on ethidium bromide (EtBr) stained agarose gels with a DNA molecular weight standard (Fermentas). Besides, both the quality and quantity of DNA were measured using the Nano Drop™ 1000 spectrophotometer as well.

Total RNA from leaves, flowers and inflorescence was isolated using RNeasy Plant Mini kit from Qiagen and Biomol kit (Biomol GmbH Hamburg). Concentration and purity of isolated RNA was determined by a standard spectrophotometer (eppendorf BioPhotometer) and Nano Drop™ 1000 spectrophotometer.

2.6 DNA sequencing and sequence analysis

For all the PCR products, plasmids and constructs DNA sequences were determined by the Max-Planck-Institute for Plant Breeding Research (MPIZ) Cologne, Germany DNA core facility on Applied Biosystems (Weiterstadt, Germany) Abi Prism 377, 3100 and 3730 sequencers using BigDye-terminator v3.1 chemistry. Premixed reagents were obtained from Applied Biosystems. Routine sequence analysis was performed using MacVector™ 7.2.3 and Assemblylign™ 1.0.9 programs. Database searches were routinely carried out using the BLAST algorithm (Altschul, 1997) at GenBank (http://www.ncbi.nlm.nih.gov) and SolGD, the Solanaceae EST database (http://www.sgn.cornell.edu).

2.7 Standard PCR conditions

The lab made Taq Polymerase and Taq DNA Polymerase (Ampliqon) were routinely used for the amplification of small DNA fragments. Large DNA fragments were amplified with LA Taq™ and PrimeSTAR™ HS (TAKARA BIO INC.), Expand High Fidelity, Expand Long Template PCR Systems (Roche) and Pfu Turbo Taq DNA Polymerase (Stratagene Europe, Amsterdam). Standard PCR reactions were performed in 25-50 µl reaction volume under the following conditions:

PCR mixture:	Cycling parameters:
5 µl 10X PCR buffer	95 °C 2 min
5 Units Taq DNA polymerase	95 °C 15 sec
1 µl 10 mM dNTP mix (dATP, dCTP, dGTP, dTTP)	Ta 30 sec
0.5 µl 10 µM sense primer	68 °C 30-90 sec
0.5 µl 10 µM antisense primer	68°C 7 min
20-50 ng genomic DNA or 1-2 ng plasmid DNA	12 °C ∞
Add H_20 to 50 µl	

n

Ta (annealing temperature of the primers) and "n" can vary for individual reactions.

2.8 Plant material

A total of 13 accessions of Withania including 5 species (*W. aristata*, *W. frutescens*, *W. riebeckii*, *W. coagulans* and *W. somnifera*) and 2 accessions of *Tubocapsicum anomalum* and one accession each for *Iochroma australe* and *Vassobia breviflora* were sampled (see Table 3.1 and 3.3). All the accessions were grown in greenhouses of the Max-Planck-Institute for Plant Breeding Research (MPIZ), Cologne. Withania and Tubocapsicum plants were grown (in the greenhouse) at 18-25°C with additional light provided during winter. For flowering, Tubocapsicum plants were grown in a climatic chamber at 22°C during the day and 16°C during the night under a 16h light/8h dark regime with constant humidity.

For transgenic plant production, *Arabidopsis thaliana* ecotype Columbia was grown under standard conditions. The seeds of *35S:AGL24* over expressing transgenic *A. thaliana* lines were kindly provided by Dr. Chaoying He (MPIZ, Cologne). Seeds were collected and stored in the seed bank of the MPIZ. For transient expression *Nicotiana benthamiana* plants were grown in greenhouses at the MPIZ, Cologne. The major plant species used in this study included: *Withania somnifera*, *Withania coagulans*, *Tubocapsicum anomalum*, and *Arabidopsis thaliana*. For all the accessions, young leaves and flowers at different developmental stages were collected, frozen in liquid nitrogen and then stored at -80°C until nucleic acid extraction was performed.

2.9 Comparison of calyx morphologies

Visual observations were taken to describe calyx morphology such as shape and architecture of the calyx in the Withaninae and closely related Iochrominae. Calyx inflation was evaluated on the basis of direct measurements of calyx length and width and on the information in

literature (D'Arcy, 1972; Zhang et al., 1994). The lengths and widths of both the flowering and fruiting calyces and fruits (10 for each) were measured using a Vernier scale. Data were analyzed statistically and accrescence and inflation of the calyx was determined.

2.9.1 Ablation experiments in Withania

Different ablation experiments were performed in *Withania somnifera*. Emasculation and removal of stigma/style was done at different stages of flower development (young bud, young flower, mature flower, young fruit, mature fruit, and old fruit) prior and after pollination. Developing fruits were removed carefully after pollination. To allow access to developing fruits, the calyx was incised longitudinally and partially removed. In addition to emasculation and fruit ablation, removal of the calyces was also done. Calyces were ablated at different developmental stages (young bud, young flower, mature flower, young fruit, mature fruit, and old fruit). In each stage at least 5 to 10 samples were ablated. Photographs were taken with a digital Nikon camera. Length and width of the calyces and fruits were measured at the time of ablation as well as after 20 days at maturity.

2.9.2 Hormone treatments

After emasculation different hormone solutions such as 6-benzylaminopurine (6-BAP), 2-isopentenyladenine (2-IP), zeatin, gibberellic acid-3 (GA3) and indole-3-acetic acid (IAA) were tested for their effects on growth of the calyx. Water served as a mock treatment. A concentration of 10-µM was applied with sterile Q-tips to the calyces of flower buds (from which styles and stigma had been removed) or flowers prior to pollination. The hormone treatment was applied continuously for 7 days. In some cases a single treatment of the hormone was given to calyces. After 20 days length and width of the calyces (5-10 for each treatment) were measured and data were analyzed. Photographs were taken with a digital Nikon camera.

2.10 Somatic chromosome count and ploidy evaluation

Withania somnifera, *W. coagulans* and *T. anomalum* were selected for the ploidy level evaluation. Their somatic chromosomes were studied in root tip meristems of the young seedlings grown on MS (Murashige and Skoog, 1962) medium with 1% agar according to Zhang et al., (2006) with modifications. Root tips approximately 1-1.5 cm in length were harvested and treated with 0.1% colchicine solution for 3 hours at room temperature. After fixing in 95% ethanol:HAc (3:1) at 4°C for at least 30 minutes or overnight, they were

preserved in 70% ethanol solution until use. In order to make them soft, the root tips were treated with 5M HCl (T001 and W007) for 15 minutes at 60°C or with 1M HCl (W002) for 10 minutes at 60°C and stained using the carbol fuchsin solution for 1 hour before squashing. The stained root tips were applied on glass slides and observed under the light microscope (ZEISS, Axiophot). At least 15-30 cells were observed to determine the number of chromosomes in each species.

2.11 Single strand cDNA synthesis

The cDNA was prepared from the total RNA. It was synthesized using SuperscriptII plus reverse transcriptase (Invitrogen). The reaction for 20 µl volume was performed as follows:

2 µg total RNA

1 µl 10 µM primer oligo$(dT)_{20}$

Add RNase/DNase free water to 12 µl.

Incubate at 70°C for 10 minutes and directly place on ice for 2 minutes.

Add the following to the tube on ice:

4 µl of 5X First Strand buffer

2 µl 20 mMDTT

1 µl 10 mM dNTP mix

1 µl SuperscriptII 2.5units/µl

Incubate for 1 hour at 42°C. After stopping the enzyme activity by incubating at 70°C for 10 minutes, precipitate in ethanol. Finally, dissolve in 100 µl TE and store at –20°C.

2.12 Gene isolation and sequencing

Chloroplast *atpB* and *matK* marker genes were amplified from genomic DNA of all the Withania and Tubocapsicum accessions using specific primers. For *MPF1-like* and *MPF2-like* genes isolation, total RNA from leaves and flowers was isolated and turned into cDNA following a standard protocol (see previous section 2.11). Both *MPF1-like* and *MPF2-like* cDNA sequences were isolated with gene primers together with the poly (T)-adaptor primers according to Clontech user manual. The sequences of all the primers used are listed in the Suppl. Table 1. The amplified fragments were cloned into pGEMT-Easy vector for sequencing. At least 12 clones for each of the gene and for each accession were sequenced from the both directions. During the isolation of *MPF2-like* genes, several pseudogenes containing pre-mature stop codons were also detected (see Table 3.3), but none of these pseudogenes were included in further analyses except MPF1-like directional evolution test.

Genomic loci were obtained with Expand Long Template PCR System (Roche) by using the gene-specific primers based on full-length cDNA sequences.

The 5'-untranslated regions (UTRs) of the Withania and Tubocapsicum *MPF2-like* cDNAs were isolated by rapid amplification of cDNA ends (RACE) using the 5'/3' RACE Kit (Roche Diagnostics, Mannheim, Germany). Gene-specific primers were designed on the basis of the cDNA sequences obtained from 3'-RACE. Putative transcription initiation sites were deduced from the sequence of the longest cDNA fragment obtained by 5'-RACE.

In addition to *MPF1-like* and *MPF2-like* gene sequences, *MAGO NASHI-like* genes were isolated from Withania and Tubocapsicum as well. In *Physalis floridana*, two copies of this gene exist, which are *PFMAGO1* and *PFMAGO2* (He et al., 2007). Therefore, primers based on the coding sequences of *PFMAGO1* and *PFMAGO2* were designed. PCR products amplified from the cDNAs of *W. somnifera* (W006) and *T. anomalum* (T001) were cloned into the pGEMT-Easy vector and plasmids were sequenced.

2.12.1 Phylogenetic tree reconstruction

Phylogenetic analyses were done in collaboration with Dr. Jinyong Hu (former group member). All the individual sequences from different clones of one gene from each accession were first edited according to original sequencing trace file and assembled using AssemblyLIGN™ (Oxford Molecular Group Plc.). The consensus sequence for each of the gene from each accession was derived according to sequence alignment of all the clones. Deduced amino acid sequences were aligned using the ClustalW method in MacVector® package (Accelrys Inc.) and hand corrected according to the alignment results using MUSCLE (Edgar and Robert, 2004) and POA (Lee at al., 2002). The nucleotide sequences were manually aligned according to the amino acid alignments. Sequences of *MPF1-like* genes were assembled to yield an alignment lacking 10 amino acids at their N-terminus. Alignments of MPF2-like proteins lacked the first 44 amino acids at the N-terminus and 3 amino acids at their C-terminus. A simple neighbour joining tree was constructed using two chloroplast markers (*atpB and matK*) and two nuclear genes (*MPF1-* and *MPF2-like-B*). To evaluate the robustness of the tree structure 1000 replicates of bootstrap searches were performed.

Phylogenetic networks could better reveal the evolutionary history including hybridization, recombination and homoplasmy etc. than a tree like structure (Huson and Bryant, 2006). Therefore, a neighbor-net network (Bryant and Moultron, 2004) reconstruction analysis was

implemented in SplitsTree4 package (Huson and Bryant, 2006) with default parameters using an uncorrected P distance method.

2.12.2 Evolutionary analyses and tests of natural selection

To examine whether the ICS undergoes an adaptive evolution, phylogeny based statistical test of positive selection on MPF1-like and MPF2-like proteins was implemented. The purpose was to determine whether any MPF1-like or MPF2-like proteins were evolved under positive selection. Because such selection may likely to act on proteins involved in morphological novelties for better fit to adaptive evolution.

For Darwinian selection tests, full length *MPF2-like* cDNA sequences were assembled from the five Withania species, from *W. aristata* the *MPF2-like-A* genes *WAA201-1*, *WAA201-2* and the *MPF2-like-B* genes *WAB201-1, WAB201-2*; from *W. coagulans WCA201-1* was excluded, since it lacked a part of its N-terminal sequence, *WCA201-2* and *WCB201-1, WCB201-2*; *W. riebeckii WRA204-1, WRA204-2* and *WRB204-1, WRB204-2*; *W. somnifera WSA206-1, WSA206-2* and *WSB206-1, WSB206-2* and *W. frutescens WFA213-1, WFA213-2* and *WFB213-1, WFB213-2*; one Tubocapsicum species, the *MPF2-like-B* gene from *T. anomalum TAB201-1*, *TAB201-2*. from the Physalis species, P068, P075, P105 and *MPF2* from *Physalis floridana*, I001 from *Iochroma australe*, Du001 from *Dunalia fasciculata*, V001 from *Vassobia breviflora,* S032 from *Solanum tuberosum* and Sm001 from *Solanum macrocarpon* (see Suppl. Table1 in Hu and Saedler, 2007). However, sequences of the 4th exon of the C-terminal region could not be aligned and hence they were removed from *MPF2-like* gene matrix. Similar dataset of MPF1-like sequences was also assembled for selection test (Table 3.3).

Maximum likelihood (ML) tree reconstruction was carried out with PAUP 4.0b10 according to Hu and Saedler (2007). To evaluate the robustness of the tree structure 1000 replicates of bootstrap searches were performed using maximum parsimony (MP) and maximum likelihood (ML) in PAUP and Bayesian posterior probability was calculated according to Ronquist and Huelsenbeck (2003). These analyses were carried out with the codeml program within PAML (phylogenetic analysis by maximum likelihood) package (Yang and Nielsen, 1998), which uses maximum likelihood estimation of various parameters to test for positive selection and to infer amino acid sites under positive selection. The analysis consists of two major steps. The first step uses the likelihood ratio test for positive selection. The ratio of ω = dN/dS with dS being the number of synonymous substitutions per synonymous site and dN the number of non-synonymous replacements per non-synonymous site was used to assess

whether the protein coding MPF1-like and MPF2-like sequences evolved under positive Darwinian selection (Yang and Nielsen, 2002). Differences between the free-ratio and one-ratio models were compared in order to check if the dN/dS ratios vary among different lineages of the phylogenetic trees. The ratio ω measures the magnitude and direction of selective pressure on a protein, with ω = 1, <1, and >1 indicating neutral evolution, purifying selection, and positive diversifying selection, respectively. The major advantage of these likelihood models over previous analysis is that they account for variable selective pressures among sites by assuming that there are different classes of sites in the proteins with different ω ratios. The second step of the analysis was to identify the amino acid residues under positive selection when likelihood ratio suggests their presence. This was achieved by calculating the posterior probabilities at each site according to data generated from different ω classes of proteins. The sites of amino acid residues with high probability of coming from a protein with ω > 1 are likely to be under positive selection.

2.13 Southern blot analysis

Southern hybridization was performed with the cooperation of Simone Riss and Dr. Chaoying He (group members). The high quality genomic DNA of *W. coagulans* (W002), *W. somnifera* (W006 and W007) and *T. anomalum* (T001) was digested with EcoRI, BamHI and KpnI restriction enzymes. The DNA fragments were separated on an agarose gel and alkaline blotted to the Hybond N^+ nylon membranes (Amersham) according to Koetsier et al., (1993). Two gene specific probes (around 300bp) for *MPF2-like-A* and *MPF2-like-B* genes were designed comprising C-terminal region and 3'UTR of these genes. For radioactive labelling of the DNA probes, 50 ng of DNA was boiled in 20 µl of H_20 for 5 minutes and quickly chilled on ice. Random oligonucleotide-primed synthesis was carried out by adding 5 µl [α^{32} P]-dCTP (10 µCi/mol), 3 µl 10X oligo buffer and 1.6 µl Klenow (2U/µl) into the probe DNA and incubating at room temperature for 1-2 hours. The Klenow enzyme was used for the fill-in reaction. Finally, the labelling product was purified using a NucleoSpin®ExtractII column.
The blot was first pre-hybridized in the hybridization buffer (Table 2.3) for 3 hours and then again hybridized in the same buffer containing radioactive labelled probe overnight at 67°C. The membrane was rinsed twice in the washing buffer and washed twice at 65°C for 20 minutes. Washed filters were exposed to a Storage Phosphor Screen (Molecular Dynamics), and signals were quantified with a Typhoon 8600 Phosphor Imager (Amersham Biosciences). Lastly, stripping was performed at 80°C in the stripping buffer and utilised for further hybridization.

Hybridization buffer (1l):	Washing buffer (1l):	Stripping buffer (1l):
150 ml 20X SSE	25 ml 20X SSE	5 ml 20X SSE
200 mg PVP 90	10 ml 10% SDS	10 ml 10% SDS
200 mg Ficoll 400	add H_20 to 1 litre	add H_20 to 1 litre
10 ml 10% SDS		
add H_20 to 1 liter		

Table 2.3: Buffers used for hybridization

2.14 *MPF2-like* gene expression analysis

Relative quantitative RT-PCR was carried out to discern the expression patterns of *MPF2-like* genes in different tissues of Withania (W002 and W006), Tubocapsicum (T002), Vassobia (V001), Iochroma (I001) and Dunalia (D001). Gene specific primers for all the accessions spanning at least 3 introns within the K-domain and the C-terminal regions were specifically designed for *MPF2-like-A* and *MPF2-like-B* genes separately. The total RNA from young leaves shorter than 1cm, mature leaves, early developing buds before the egression of petals, buds before anthesis, and flowers at anthesis was extracted with RNeasy Plant Mini Kit (Qiagen) and treated with DNase I (Roche) to remove the genomic DNA contamination. The first strand cDNA synthesis was carried out using the Clontech PowerScript II reverse transcriptase with 2 µg of total RNA in a 20µl reaction volume. Normalization was done using constitutively expressed *18S* rRNA. Multiplex RT-PCR was carried out in 20 µl reaction volumes using Qiagen Taq polymerase for *MPF2-like* primer pairs for 14 cycles and continued to another 16 cycles after the inclusion of 18S primers. Amplified PCR products were run on 2% agarose gel and photographed. The intensity of amplification (*MPF2-like* expression) was judged after scanning the gel on Typhoon 8600 Phosphor Imager (Amersham Biosciences). All the fragments were sequenced and proved to be identical to the *MPF2-like* sequences, isolated before. The experiment was repeated 3 times for each of the 2 independent samples.

2.15 Over-expression analysis of *MPF2-like* genes in Arabidopsis

In order to discern the functional divergence of *MPF2-like* genes, over-expression of these genes in Arabidopsis was investigated. Two types of vector backbones were used for this purpose. In the first case, three constructs, *35S:WSA206:YFP*, *35S:WSB206:YFP* and *35S:TAB201:YFP* with C-terminal YFP tag were generated by using gateway cloning technology (Invitrogen) according to the manufacturer's protocols. Full-length (sense) cDNAs

of *WSA206*, *WSB206* and *TAB201* were cloned into the pENTR/D-TOPO-201 vector and finally recombined with destination vector pDES-XCS-mYFP using clonase 2.1 (Invitrogen). The constructs were sequenced to verify that YFP gene is in correct orientation and coding frame with *MPF2-like* genes driven by the CaMV35S promoter.

In the second approach, a two-step cloning strategy was adopted. As a first step, full-length (coding sequence) cDNAs of *WSA206*, *WSB206* and *TAB201* were cloned into the pRT100 vector by using BamHI and NcoI enzymes restriction sites. In the second step, whole cassette containing insert was cleaved with HindIII enzyme and shuttled into plant binary vector pBAR-A for expression driven by CaMV35S promoter. These constructs were also verified by sequencing. Transformation of all the constructs without YFP and with YFP tag along with empty vectors into *Arabidopsis thaliana* was mediated by GV3101 (pMP90RG) and GV3101 (pMP90RK) strains of *Agrobacterium tumefaciens,* respectively. For transformation floral-dipping protocol (Clough and Bent, 1998) was used with major modifications such that MS media (Murashige and Skoog salt mixture) and B5 (Gamborg) and vitamins were excluded from the agro-infiltration media. After germination, the transgenic plants harbouring constructs were screened by spraying with 0.15-0.2% BASTA twice a week. Transgenic plants containing insert were confirmed by RT-PCR analysis using gene specific primers.

In order to observe the phenotypic differences, leaves and inflorescences of the transgenic plants expressing *MPF2-like* genes were observed visually. The pedicel length of transgenic Arabidopsis plants containing *MPF2-like* genes along with control plants was measured with Vernier scale. The flowers with fruit were photographed using a digital camera or camera attached to a binocular microscope. The images of transgenic plants were acquired at the same developmental stage (flowering stage) under the same magnification. In case of transgenic plants transformed with YFP-tagged constructs, leaves and calyces of transgenic plants were scanned under a Leica TCS SP2 AOBS^{R}, Confocal Laser Scanning Microscope (CLSM) to monitor the localization and intensity of YFP signals.

2.16 Isolation of *MPF2-like* genes promoter and first intron sequences

A total of 16 promoter and 14 first intron sequences of *MPF2-like* genes were isolated from various accessions of Withania, Tubocapsicum and Vassobia by RAGE (rapid amplification of genomic DNA ends) method (Clontech). The complete list of all the promoter and intron sequences isolated with corresponding length is given in Table 2.4. For this purpose genomic DNA was completely digested with different blunt end cutting enzymes e.g. DraI, EcoRV, ScaI, SmaI, PvuII, StuI, and XbaI over night. After purification with spin columns

(MACHEREY-NAGEL, Germany) the fragments were ligated to adapters using T4 DNA ligase enzyme (Roche) at 4°C. The first PCR was carried out using an adapter primer AP and gene specific primer, while the second PCR was executed with nested adapter primer NAP and gene specific primer, after diluting the 1st PCR product 50 times with water. Gene specific primer sequences are listed in the Suppl. Table 1. The adapters and corresponding primers are described in the manual (Clontech), and the gene-specific primers were the same as in the 5'-RACE.

MPF2-like 1st intron sequences were isolated with PCR amplification by using primers designed to bind to promoter region and the second exon of *MPF2-like* genes.

All the *MPF2-like* promoter and 1st intron sequences were isolated as a single contiguous piece and their gene identities were confirmed by comparing them with the corresponding full length cDNA sequences especially the 5'-RACE sequences.

2.16.1 *MPF2-like* promoter and first intron sequence analyses

For *MPF2-like* genes, 1.2 kb promoter sequences up-stream and 1.5 kb 1st intron sequences down stream (length of 1st intron is more than 1.8 kb in all the genes) of the ATG were selected for bioinformatic analyses. All the sequences were analyzed using programs TBA and MultiTF (http://rvista.dcode.org/cgi-bin/mTTP.cgi) in the Mulan (Multiple-sequence local alignment) Package Online (http://mulan.dcode.org). Putative *cis*-acting elements in the promoters were predicted using the plant *cis*-acting regulatory DNA elements database Transfac (http://www.gene-regulation.com/pub/databases.html).

Concatenated tba.maf files were submitted to MultiTIF in TRANSFAC professional V10 library from plants and matrix similarity optimized for functions was predefined as 0.85. Eleven out of 80 plant transcription factor families were selected for the creation of TBX5 and NKX 2.5 high quality matrix. The parameters selected for the dynamic overlay of transcription factor binding sites (TFBS) prediction with conservation profile clustering included: 3 kb per layer, picture width of 800 pixels, one site per 1000 bp and clustering with smooth plot. Finally, the dynamic visualization of TFBS was attained in a graphical format.

2.16.2 Expression analysis of *MPF2-like* promoter:GUS and promoter-1st exon1st intron:GUS constructs in transgenic Arabidopsis

Based on bioinformatic analyses from the promoter and 1st introns sequence, *WSA206*, *WSB206*, *TAB201*, *VBB201* and *MPF2* genes were selected for the generation of promoter:GUS and promoter-1st exon-1st intron:GUS constructs. A total of 39 (9 for *WSA206*;

9 for *WSB206*; 9 for *TAB201*; 5 for *V001*; and 7 for *MPF2*) deletion constructs keeping in consideration the important motifs identified for the above-mentioned genes were designed. More details about the pursued scheme of construct making is given in the Suppl. Fig. 1 and specifications of the constructs are mentioned in Table 3.5.

The GUS expression vector pGVT-bar was used as backbone for making these constructs. The promoter, intron and vector were digested with SmaI, XmaI, HindIII, SbfI and XbaI restriction enzymes. After purification, ligation was done using T4 DNA ligase from NEB (New England Biolabs) at 15 ^{o}C overnight. The constructs were sequenced to confirm the correct frame and orientation of the insert.

Sr. Nr.	Accession	Species	Genes	Promoter length (bp)	Intron length (bp)
1	W001	*W. aristata*	*WAA201*	2667	2373
2	W001	*W. aristata*	*WAB201*	3294	1827
3	W002	*W. coagulans*	*WCA202*	2609	2541
4	W002	*W. coagulans*	*WCB202*	2978	1865
5	W004	*W. riebeckii*	*WRA204*	2645	2621
6	W004	*W. riebeckii*	*WRB204*	3015	1873
7	W006	*W. somnifera*	*WSA206*	2645	2589
8	W006	*W. somnifera*	*WSB206*	2919	1870
9	W007	*W. somnifera*	*WSA207*	2606	2479
10	W007	*W. somnifera*	*WSB207*	3348	1945
11	W013	*W. frutescens*	*WFA213*	2645	2452
12	W013	*W. frutescens*	*WFB213*	2970	2001
13	T001	*T. anomalum*	*TAB201*	2308	3704
14	T002	*T. anomalum*	*TAB202*	2357	-
15	V001	*V. breviflora*	*V001_1*	2107	-
16	V001	*V. breviflora*	*V001_2*	2056	3386

Table 2.4: List of promoter and first intron sequences with corresponding length

Transformation of all the constructs along with empty vectors into Arabidopsis was mediated by GV3101 (pMP90RG) strain of *A. tumefaciens*. For plant transformation, as previously mentioned, the floral-dipping protocol (Clough and Bent, 1998) was used with major modifications such that MS media (Murashige and Skoog salt mixture) and B5 (Gamborg) and vitamins were omitted from the agro-infiltration media.

Histochemical GUS analysis was carried out to observe the expression patterns of *MPF2-like* genes in different tissues of the transgenic Arabidopsis plants according to Jefferson et al., (1987). Inflorescence and rosette leaves of the transgenic plants were harvested and immediately sub-merged into the GUS staining solution. The samples were vacuumed for 15-30 minutes twice to allow the GUS solution to penetrate into the tissues and incubated at 37^{o}C overnight. The samples were made free of chlorophyll by treating them with 70% ethanol at least three times. Finally, GUS expression patterns were observed visually as well as with the binocular.

2.16.3 Expression analysis of *MPF2-like* promoter:YFP and promoter 1st exon-1st intron:YFP constructs in transiently transformed *Nicotiana benthamiana* leaves

The gateway expression vector PXCG-mYFP without promoter was used for making YFP constructs. For generating YFP constructs, exactly the same *MPF2-like* promoter and intron sequences as described in previous section (2.16.2) for GUS analysis, were used (see Suppl. Fig. 1 and Table 3.5). A total of 36 constructs (9 for *WSA206*, 9 for *WSB206*, 9 for *TAB201*, 9 for *MPF2*) were generated using the gateway cloning technology (Invitrogen) according to the manufacturer's instructions. The promoter and intron sequences were cloned into pENTR/D-TOPO-201 vector and finally recombined with destination vector pDES-XCG-mYFP using clonase 2.1 (Invitrogen). After sequencing the plasmids containing inserts in the correct orientation and coding frame were transformed into the *A. tumefaciens* strain GV3101 (pMP90RK) for transient infiltration.

For transient infiltration liquid YEB media containing appropriate antibiotics was inoculated with GV3101 (pMP90RK) harbouring promoter and intron constructs as well as P19 proteins constructs. At required OD cells from both the cultures were harvested and re-suspended in the infiltration buffer containing acetosyringone. The P19 is one of a class of plant and animal virus proteins that are known to interfere with defence mechanism of the host (Molnar et al., 2005). The $OD_{600:}$ 2 was set for all the constructs cultures as well as for the P19 protein construct. The cultures were mixed before infiltration into *N. benthamiana* leaves as described by Llave et al., (2002). The upper two healthy leaves were co-infiltrated with a 1ml plastic syringe.

Three days after infiltration, leaves of *N. benthamiana* were scanned with Leica TCS SP2 AOBS^{R}, Confocal Laser Scanning Microscope (LSCM) for YFP signal detection. At least 10 images showing YFP signals in the nuclei were selected randomly to quantify the luminescence with ImageJ program. The intensity of YFP signals with a particular construct

was considered as the strength of the *MPF2-like* gene promoter and intron. Data were processed and analyzed to reveal the strength of the promoters and introns.

2.17 Yeast two-hybrid analyses

For detecting MPF2-like and MAGO NASHI-like protein-protein interactions in the yeast system, two yeast strains AH109 and Y187 as well as two yeast vectors pGBKT7 (bait) and pGADT7 (prey) were employed (Clontech). Preparation of yeast media, competent cells, co-transformation and bait-prey mating was done according to the methods described in the yeast protocols handbook (Clontech).

2.17.1 Generation of MPF2-like and MAGO NASHI-like proteins constructs

A total of 17 yeast constructs for MPF2-like and MAGO NASHI-like proteins from Withania, Tubocapsicum, Physalis and Arabidopsis were tested in this study (Table 2.5). The nine constructs were generated in this study while, eight constructs were provided by Dr. Chaoying He (MPIZ, Cologne). All the bait and prey constructs used in this study are listed in Table 2.5. Full-length cDNAs of WSA206, WSB206, TAB201, WSMAGO1, WSMAGO2, TAMAGO1 and TAMAGO2 were digested with NcoI and BamHI restriction enzymes. After purification with spin columns, digested vectors (pGBKT7 and pGADT7) and coding sequences were ligated using T4DNA ligase (NEB) at 4 ^{o}C for two overnights. For verification, all the constructs in both the vectors were sequenced.

In addition, WSA206ΔC246 and WSB206∇C257 yeast constructs were also generated by using the same methodology as described above.

2.17.2 Tests for auto-activation, homo- and heterodimer formation and interactions

Auto-activation and auto-binding was tested by co-transformation of MPF2-like and MAGO NASHI-like proteins with empty bait and prey vectors. No growth of the transformants on SD/-Trp-Leu-His solid plates indicated that all the MPF2-like and MAGO NASHI-like proteins could not activate the HIS3 reporter gene on its own i.e. could not self-activate and self-bind. Thus, the full-length version (pGBKT7-MPF2-like and pGBKT7-MAGO NASHI-like) could be used as bait not only to screen 2-hybrid cDNA libraries and but to test their interactions as well. Furthermore, their homo- and hetero-dimerization potential was tested. In addition, WSA206ΔC246 and WSB206∇C257 were also tested for their auto-activation, auto-binding and dimerization properties. Finally all the constructs were co-transformed to in yeast strain AH109 to study their interactions with each other. Both medium-stringency (SD/-Trp-

Leu-His) and high-stringency (SD/-Trp-Leu-His-Ade) yeast media were used for plating. The yeast plates were incubated at 28 °C for 2-4 days. Colonies bigger than 2 mm were dissolved in medium-stringency liquid media (SD/-Trp-Leu-His) and incubated at 28 °C for 2 overnights. Finally, the interaction matrix was generated on both the medium-stringency (SD/-Trp-Leu-His) and high-stringency (SD/-Trp-Leu-His-Ade) yeast media plates.

2.17.3 Screening of Arabidopsis oligo-dT cDNA yeast library with MPF2-like proteins

Arabidopsis oligo-dT cDNA yeast two-hybrid library was kindly provided by Markus Kuckenberg (MPIZ, Cologne). Three pGBKT7-MPF2-like bait constructs (WSA206, WSB206 and TAB201) were transformed into the Y187 yeast strain. The oligo-dT cDNA library in the AH109 strain was served as a prey. Yeast mating was done according to library screening protocols (Clontech). For yeast colonies growth, high-stringency solid media (SD/-Trp-His-Leu-Ade and 3.0 mM 3-AT) were used. Yeast cells were incubated in a growth chamber at 28 °C for 4-7 days. Yeast manipulations were performed following standard procedures (Clontech). Colonies bigger than 2mm were treated with the lyticase to dissolve the walls and standard PCR was performed to amplify the interactors using primer pair designed from prey vector sequence as listed in the Suppl. Table 1. PCR products were cleaned with ExoSAP-IT PCR Clean-up Kit (USB Corporation) and sequenced. The sequences were subjected to BLAST (Basic Local Alignment Search Tool) search using blastx program (Search protein database using a translated nucleotide query) in NCBI (National Centre for Biotechnology Information) database. A list of interactors with number of hits was generated for all the three baits (WSA206, WSB206 and TAB201) separately.

To verify the interactions, SEP1, SEP4 and SOC1 bait constructs were co-transformed in AH109 yeast strain. In addition, other constructs like AP1, SEP2, and SEP3 were also tested for interactions. Transformed mixture of these cells was plated on medium-stringency (SD/-Trp-His-Leu) and high-stringency plates (SD/-Trp-His-Leu-Ade and 3.0 mM 3-AT), respectively.

The non-lethal ß-galactosidase assay was performed as described by Duttweiler (1996) to confirm all above-mentioned interactions detected.

Sr. Nr.	Plant species	Protein	pGBKT7 (Bait)	pGADT7 (Prey)	Source
1	*W. somnifera*	WSA206	WSA206	WSA206	This study
2	*W. somnifera*	WSB206	WSB206	WSB206	This study
3	*W. somnifera*	WSA206	WSA206ΔC246	WSA206ΔC246	This study
4	*W. somnifera*	WSB206	WSB206∇C257	WSB206∇C257	This study
5	*W. somnifera*	WSMAGO1	WSMAGO1	WSMAGO1	This study
6	*W. somnifera*	WSMAGO2	WSMAGO2	WSMAGO2	This study
7	*T. anomalum*	TAB201	TAB201	TAB201	This study
8	*T. anomalum*	TAMAGO1	TAMAGO1	TAMAGO1	This study
9	*T. anomalum*	TAMAGO2	TAMAGO2	TAMAGO2	This study
10	*P. floidana*	PFMAGO1	PFMAGO1	PFMAGO1	Chaoying He
11	*P. floridana*	PFMAGO2	PFMAGO2	PFMAGO2	Chaoying He
12	*A. thaliana*	AP1	AP1ΔC196	AP1	Chaoying He
13	*A. thaliana*	SEP1	SEP1ΔC168	SEP1	Chaoying He
14	*A. thaliana*	SEP2	SEP2ΔC168	SEP2	Chaoying He
15	*A. thaliana*	SEP3	SEP3ΔC171	SEP3	Chaoying He
16	*A. thaliana*	SEP4	SEP4	SEP4	Chaoying He
17	*A. thaliana*	SOC1	SOC1	-	Chaoying He

Table 2.5: List of constructs used in yeast two-hybrid analysis

3 Results

The results part has been structured into four sections. In the first section, calyx morphology, ablation of flower organs and particularly, calyx and fruit growth in response to hormones application have been elucidated. Second section, which is the major part, is mainly concerned with phylogenetic analyses of chloroplasts markers, nuclear trait determining genes such as *MPF1-like* and *MPF2-like*, their duplication, diversification and expression in different tissues. Additionally, functional characterization of *cis*-elements regulating the *MPF2-like* genes expression and their protein-protein interactions has been illustrated in the 3rd and 4th sections, respectively.

3.1 Calyx morphology, architecture and effects of hormones on calyx growth

Several of the nine Withaninae genera feature ICS including Mellissia, Nothocestrum, Physaliastrum and Withania, while Tubocapsicum, for example, only has a small inconspicuous calyx (D'Arcy, 1991). In the following, closely related Withania and Tubocapsicum species were taken as examples for ICS and non-ICS featuring species. A total of 13 accessions of Withania including 5 species (*W. aristata*, *W. frutescens*, *W. riebeckii*, *W. coagulans* and *W. somnifera*) and 2 accessions of *Tubocapsicum anomalum* and one accession each for *Iochroma australe* and *Vassobia breviflora* were used in the study (Table 3.1 and 3.3) with a special emphasis on *W. somnifera*.

3.1.1 Calyx diversity in the Withaninae

Variations in calyx shape and architecture are visible in the Withaninae. While sepals of *W. coagulans* and *W. somnifera* are completely fused and encapsulate the berry, the sepals of *W. riebeckii*, *W. aristata* and *W. frutescens* are not fused at the apex and constitute an open lantern (Fig. 3.1A,B,C,D,E,F). The half balloon of *W. aristata* contains needle-like projections at the tip unlike *W. frutescens*, which shows teeth-like invaginations (Fig. 3.1A,E). *W. somnifera* has a more globose lantern as compared to *W. coagulans*, where it is longer in length than width at the apex, especially. Both *W. somnifera* and *W. coagulans* have semi-succulent and papery calyces, while more fleshy sepals are common in other species such as *W. riebeckii*. While Withania displays a diversity of ICS, *Tubocapsicum anomalum* is completely devoid of a true calyx and features only a rudimentary calyx. During fruit development, the flowering and fruiting calyces of Withania can change in size and/or architecture.

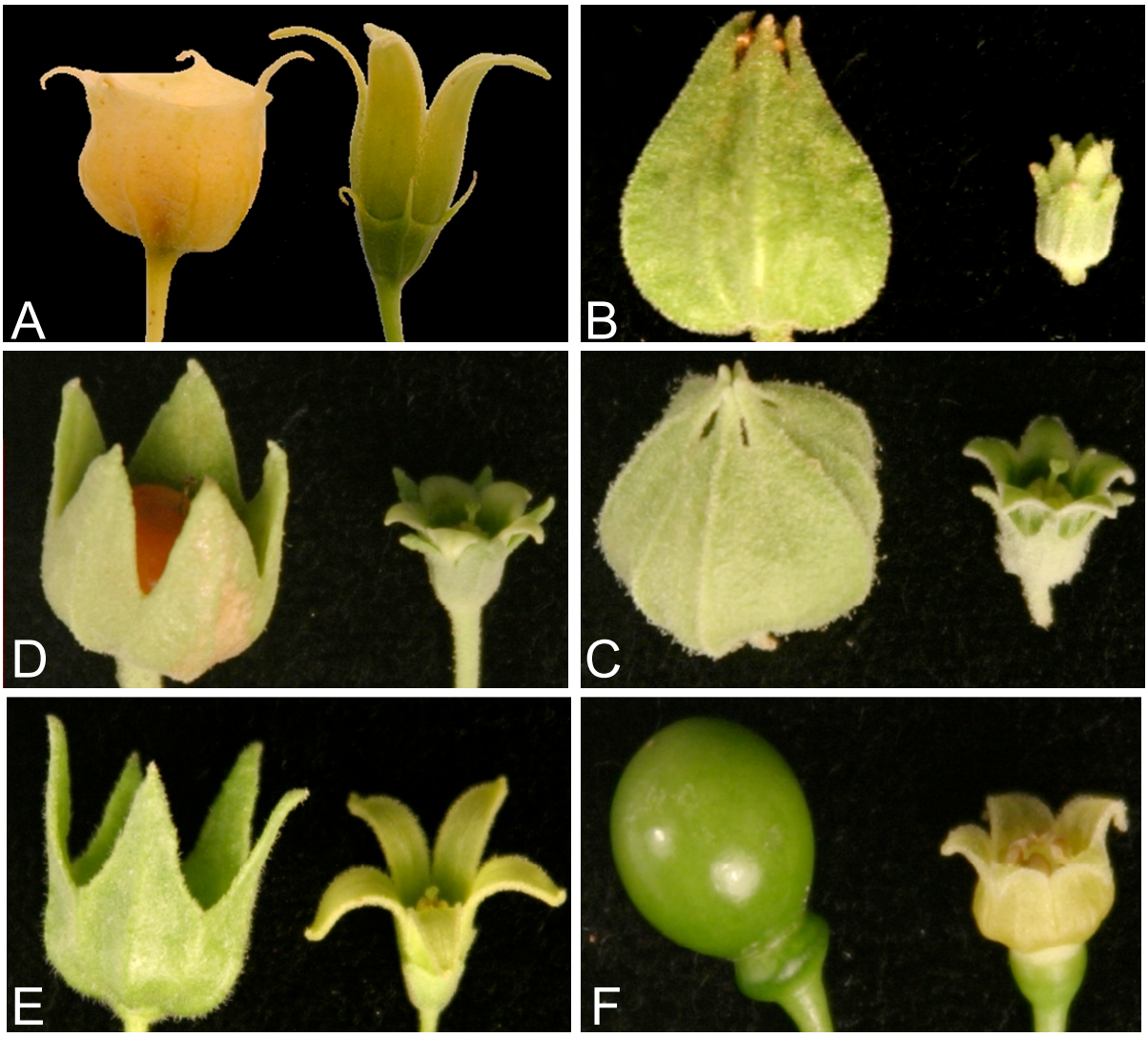


Figure 3.1: Calyx diversity in Withania and Tubocapsicum

Photographs exhibiting the diversity in flowering and fruiting calyces of Withania and Tubocapsicum are shown. A) *Withania aristata*; B) *Withania coagulans*; C) *Withania riebeckii*; D) *Withania somnifera*; E) *Withania frutescens*; F) *Tubocapsicum anomalum.* Note rudimentary calyx in Tubocapsicum.

3.1.2 Variations in the inflation degree of the calyces

Lengths and widths of the flowering and fruiting calyces were measured in various accessions to quantify the enlargement of the calyx during fruit maturation (accrescence), and data were compared with reports from the literature (D'Arcy, 1973; Zhang et al., 1994; Hu and Saedler, 2007) including the visual impressions of calyx architecture described. Table 3.1 shows that in Withania the average increase in fruiting calyx is about 2 to 3 times. *W. somnifera* (W006, W007, W008, W012) and Withania species (W010 and W011) demonstrate the maximum increase in the calyx growth. W010 exhibits the highest value both in calyx length ratio (3.83)

and calyx width ratio (4.46). Among the accessions showing least values of calyx inflation comprise W001 (*W. aristata*) and W013 (*W. frutescens*). *W. riebeckii* (W004 and W005) manifests an intermediate increase in the calyx inflation. Generally, the tendency of increase in calyx size is more in width than length in all the Withania accessions studied. Contrary to Withania, there is little increase (length ratio, 0.94 and width ratio 0.75) in fruiting calyx of Tubocapsicum. Both the flowering and fruiting calyces are almost of the same sizes. Iochroma and Vassobia also show no calyx growth during fruit maturation. Only Withania features accrescence and inflation in the calyx as evident from the ratios observed.

				Calyx morphology							
				Calyx length (mm)			Calyx width (mm)				
Accession	Genus	Species	Source	Flower	Fruit	Ratio (Fr/Fl)	Flower	Fruit	Ratio (Fr/Fl)	Accrescence	Inflation
W001	*Withania*	*aristata*	BGN	07.4±01.1	16.8±01.5	2.27	04.2±00.8	10.6±01.2	2.54	+	+
W002	*Withania*	*coagulans*	BGN	04.6±00.6	14.9±02.2	3.24	02.8±00.7	10,9±00.8	3.89	+	+
W003	*Withania*	*frutescens*	BGN	04.0±00.4	14.5±00.8	3.62	03.0±00.4	12.5±02.0	4.17	+	+
W004	*Withania*	*riebeckii*	BGN	05.8±01.4	10.2±00.3	1.76	03.8±01.4	07.8±00.1	2.05	+	+
W005	*Withania*	*riebeckii*	BGN	05.6±01.1	11.5±00.9	2.05	04.0±00.6	08.6±00.4	2.15	+	+
W006	*Withania*	*somnifera*	BGN	06.4±01.7	16.8±01.3	2.62	03.7±01.4	12.3±03.1	3.32	+	+
W007	*Withania*	*somnifera*	BGN	03.5±00.6	13.4±00.6	3.83	02.8±00.8	11.3±00.8	4.03	+	+
W008	*Withania*	*somnifera*	BGN	04.8±00.9	15.7±02.1	3.27	03.2±02.0	12.8±00.6	4.00	+	+
W009	*Withania*	*ssp.*	BGN	04.7±00.8	15.6±00.1	3.32	02.9±00.3	09.8±0.13	3.37	+	+
W010	*Withania*	*ssp.*	BGN	03.6±01.0	13.8±03.0	3.83	02.8±00.9	12.5±00.5	4.46	+	+
W011	*Withania*	*ssp.*	BGN	04.0±00.8	13.6±00.7	3.40	03.3±01.1	14.6±01.3	4.42	+	+
W012	*Withania*	*somnifera*	BGN	03.9±00.5	12.0±02.0	3.08	02.6±00.3	11.6±00.4	4.46	+	+
W013	*Withania*	*frutescens*	UPU	07.3±00.4	14.3±00.3	1.96	05.6±00.7	10.7±00.9	1.91	+	+
T001	*Tubocapsicum*	*anomalum*	BGN	02.1±00.3	02.0±00.0	0.94	02.4±00.1	03.2±00.2	0.75	-	-
T002	*Tubocapsicum*	*anomalum*	BGN	02.3±00.2	02.2±00.1	0.96	02.7±00.3	03.3±00.6	1.22	-	-
I001	*Iochroma*	*australe*	BGN	07.6±004	09.7±0.9	1.28	06.6±0.8	08.6±01.1	1.30	-	-
V001	*Vassobia*	*breviflora*	BGN	03.4±00.9	03.8±01.1	1.12	03.1±0.02	03.9±00.7	1.23	-	-

Table 3.1: List of accessions used for calyx measurements

All the accessions listed were grown in the glasshouses of MPIZ Cologne and verified for their ICS phenotype. Source shows seed bank providing the seeds. BGN, Botanical and Experimental Garden of Radbound University, Nijmegen, the Netherlands; UPU, Uppsala University Sweden.

Length and width of 10 calyces were measured using a Vernier scale and standard deviations were determined. Accrescence was determined by visual observation of calyx expansion in fruiting calyx; +, accrescence; -, normal. Inflation means accrescent calyces (almost) completely encapsulating the fruit; +, inflation (Inflated calyx syndrome, ICS); -, normal.

From the above observations, it can be inferred that there is definitely an increase in the flowering calyx size after pollination in Withania. The assumption can be the production of a signal which, after pollination induces the growth response in calyx. Obviously, some hormones might play a role in this connection. But before hormone applications, it is

imperative to know that calyx growth occurs after pollination. This can be analyzed by ablation experiments in Withania.

3.1.3 ICS develops after pollination and crosstalk occurs between the calyx and fruit

The ablation experiments were done with *W. somnifera* (W007). When stigma of the young bud was removed, flower dried up and decreased in size, which ultimately, resulted in the reduction of calyx size, especially the width (Fig. 3.2 A,B). When removal was done in young flower, the calyx grew albeit in size, in length particularly, then stopped the growth and flower senescenced earlier. Removal of the stigma in pollinated flowers led to a development exactly like in normal control flower and fruit production. This confirmed that pollination has surely occurred before the removal of stigma at this stage. There was no growth of calyx after removing the fruit at young and old stages, although sepals became more elongated than broader in case of young fruit.

On the other hand, removing sepals at young bud or flower stage completely abolished the fruit setting. The flower dried up and decreased in size. Calyx ablation at pollinated flower stage resulted in the normal fruit development and flower remained persistent. Removing the calyx at later stages always maintained the fruit development (Fig. 3.2 C,D). When calyx was removed from top, some of the basal calyx still remained attached to fruit. Basal calyx was able to elicit the normal fruit development. Similar experiments were done with Tubocapsicum, but removal of either stigma or fruit always resulted in the senescence of the flower (data not shown).

Ablation experiments involving the removal of stigma, sepals or developing fruits suggested that crosstalk occurred between the sepals and developing fruit during ICS formation. Preventing fertilization by emasculation or by removal of stigma/style at the flower-bud stage inhibited further calyx formation, while ablation of fertilized carpel soon after pollination abrogated the sepal growth and therefore prevented ICS formation. However, if the developing berry was removed after pollination, ICS formation still occurred, although resulting structure was smaller than wild type. This implies that pollination or early steps in fertilization provide substances either signals or nutrients for the sepal growth. These results are in agreement with He and Saedler (2007). On the other hand, ablation of sepals before pollination prevented the fruit development, while ablations done at later stages allowed the development of smaller berries, suggesting that sepals also contribute to fruit development, as in the case of *Helleborus foetidus* and *Physalis floridana* (Herrera, 2005; He and Saedler, 2007).

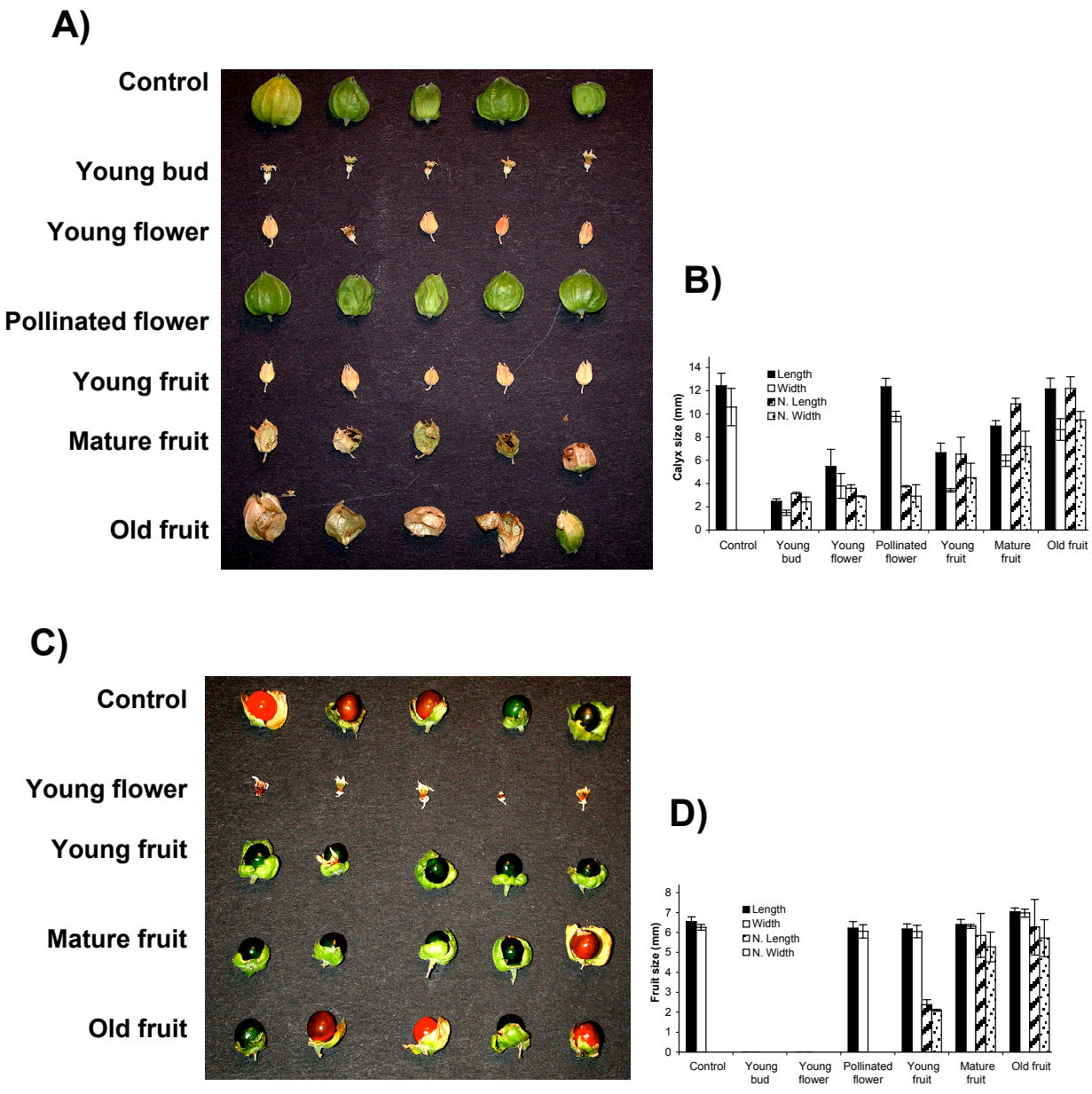


Figure 3.2: Crosstalk between calyx and developing fruit as revealed by ablation experiments in *Withania somnifera*

A) Photograph showing the calyx growth after removal of stigma at different stages (young bud, young flower, pollinated flower and fruit at young, mature and old stages). Note the growth of sepals at pollinated flower stage and no growth at young bud stage. B) Graph showing variations in sepal size after removal of stigma and fruit at different stages. Note the increase in sepal size after stigma/style ablation at pollinated flower stage. Lengths and widths are indicated as black and white columns, respectively. Normal (N) lengths and widths (sizes) at the time of ablation are indicated as striped and spotted white columns , respectively. Error bars indicate standard deviation. C) Photograph showing the fruit growth after ablation of calyx at stages such as young flower, young fruit, mature fruit and old fruit at maturity. Note no initiation of fruit at young flower stage. D) Graph indicating the effects of calyx removal on fruit size at different stages. Compare the size of fruit after removal of calyx at young flower and pollinated flower stages. Like C) Lengths and widths are indicated as black and white

columns, respectively. Similarly, Normal (N) lengths and widths (sizes) at the time of ablation are indicated as striped and spotted white columns, respectively. Error bars indicate standard deviation.

Thus, in *W. somnifera* sepals affects the fruit development and fruit bring about the changes in sepals growth. Definitely, there is some kind of crosstalk occurring between them. It was attempted to determine the nature of these signals. Developing fruits for example, are known to produce auxins, cytokinins and gibberellins (Ozga and Reinecke, 2003; Weyers and Paterson, 2001). As a next step, role of hormones was focused and their effects on ICS development were addressed further.

3.1.4 Cytokinins and gibberellins are essential for ICS formation in Withania

From previous section it is evident that normal *W. somnifera* displays fully developed ICS, while ablation of stigma of flower buds completely prevents ICS formation. Fig. 3.3A depicts that ablated flowers become dry when treated with water or auxin. Both are unable to restore the ICS formation in Withania. In contrast, treatment of stigma ablated flower buds with a cytokinin solution (10-μM of 6-benzylaminopurine, 6-BAP, 2-isopentenyladenine, 2-IP, zeatin or a mixture of all) or a gibberellin solution (gibberellic acid-3, GA3) restore ICS formation. Similar results were obtained by He and Saedler (2007), while treating *Physalis floridana* with hormones. Although ICS developed in flower buds treated with gibberellic acid, the increase was more in length than width. Moreover, sepals treated with gibberellins and cytokinins remained persistent unlike in water and auxins treatments, where sepals dried out and ultimately dropped. This suggests that cytokinins not only trigger ICS but also maintain it. The tiny berries that developed in these 'lanterns', however, did not contain any seeds. Contrary to Withania, Tubocapsicum plants treated with either of the hormones showed no response and no ICS like structures developed.

3.1.5 Single exposure to cytokinins is sufficient to induce ICS in Withania

Although Withania sepals were continuously treated with cytokinins for 7 days but it seemed that a single treatment could also induce the increase in calyx length. Remarkably, the sepals became persistent after a single exposure to hormones (data not shown). A single treatment with a mixture of GA3 and 6-BAP is sufficient to induce the formation of a lantern comparable in size with those seen under the natural conditions of self-fertilization (He and Saedler, 2007).

Thus, a treatment of calyx with these hormones can mimic the effect of fertilization on calyx growth. The Withaninae are a group of wild plants. Not much is known about their ploidy level. Determination of their chromosome number will help to define the molecular basis of ICS formation in Withania and lack of ICS formation in Tubocapsicum.

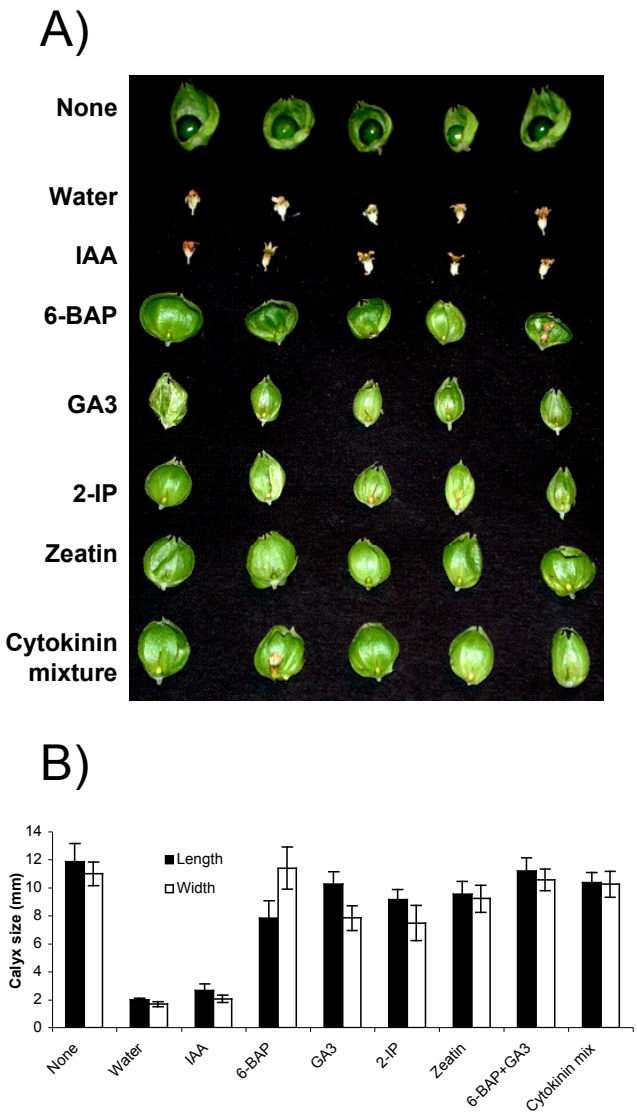


Figure 3.3: Hormonal treatment of stigma/style-ablated flower buds of *Withania somnifera*

A) Inflated calyx syndrome (ICS) phenotypes obtained following diverse regimes of treatments indicated. A concentration of 10-μM was administered. Part of the calyx was removed to expose the developing berry. Note that the use of indole-3-acetic acid (IAA) or water does not restore calyx growth and no mature berry developed in all the treated samples. B) Graph indicating average sepal lengths and widths obtained under different treatments. Lengths and widths are indicated as black and white columns, respectively. Error bars indicate standard deviation.

3.1.6 Withania and Tubocapsicum are tetraploids

The haploid (n) number of chromosomes in *Withania* and *Tubocapsicum* is 12 like in other Solanaceaous plants. An evaluation of the ploidy level of Withania and Tubocapsicum was undertaken. Among the 5 species of *Withania*, the ploidy of *W. coagulans* and *W. somnifera* has been reported to be at least 4n (Baquar, 1967; Bir and Neelam, 1980; Iqbal and Datta, 2007). *T. anomalum* also seemed to be tetraploid (D'Arcy et al., 2001). In the beginning efforts were made to determine the ploidy level in these species by flow cytometry but data generated could not be analyzed properly mainly due to lack of proper controls. Alternatively, the classical root squash experiment was performed. In this experiment at least 15 crashed cells for each species were examined. The chromosome number varied from 34 to 48 as minimum and maximum in these cells (Table 3.2). The average number came out to be 39.42±3.38 for *W. coagulans*, 39.53±3.46 for *W. somnifera* and 43.52±3.37 for *T. anomalum*. In some cells the chromosome seemed to be lying above one another and not good enough scattered to be counted easily (Fig. 3.4). The maximum number of chromosomes for *W. coagulans*, *W. somnifera* and *T. anomalum* was counted to be 48, 47, and 48, respectively. Thus counting of chromosomes on root tip squashes of *W. coagulans, W. somnifera* and *T. anomalum* verified that all above species seem to be tetraploid (4n=48) which, substantiate the already published reports.

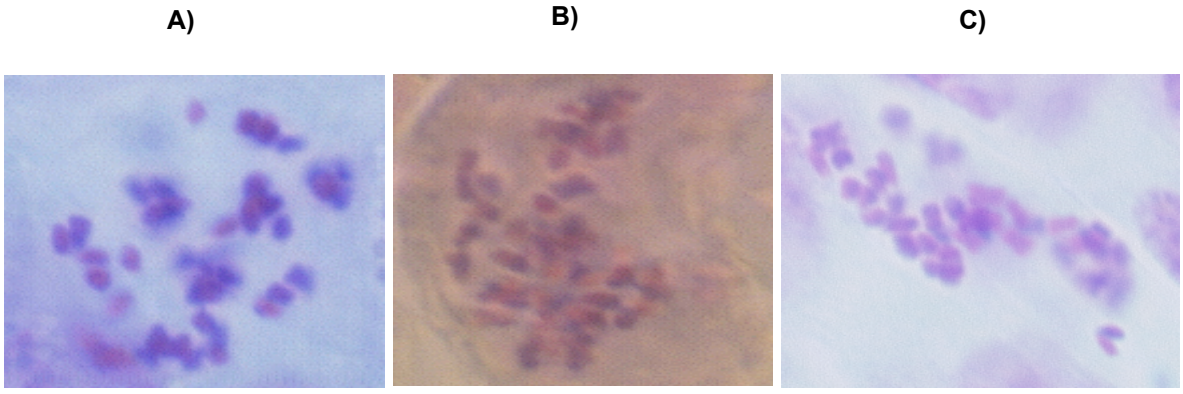


Figure 3.4: Somatic cells showing chromosomes stained with carbol fuchsin solution

A) *W. somnifera*; B) *T. anomalum*; C) *W. coagulans*

Therefore, as a next step, these results were verified on the molecular level through cDNA cloning of various genes and Southern blotting experiments.

Cell Nr.	*W. coagulans*	*W. somnifera*	*T. anomalum*
1	42	41	42
2	35	36	44
3	41	42	46
4	48	39	42
5	33	38	47
6	36	43	39
7	41	39	46
8	39	35	41
9	42	34	47
10	39	38	44
11	40	40	38
12	38	44	45
13	39	47	48
14	37	39	47
15	41	38	37
16	43		45
17	35		41
18	40		43
19	40		39
20	35		48
21	42		45
22	36		
23	36		
24	37		
25	39		
26	39		
27	36		
28	42		
29	40		
30	42		
31	48		
32	41		
33	39		
Average	**39.42**	**39.53**	**43.52**
St. Dev.	**±3.38**	**±3.46**	**±3.37**

Table 3.2: Somatic cell chromosome count from root tip squashes of *W. coagulans*, *W. somnifera* and *T. anomalum*. Red box indicates the maximum number of chromosomes counted for each species.

3.2 Duplication, selection, diversification and sub-functionalization and/or neo-functionalization of *MPF2-like* genes in the Withaninae

In diploid *Physalis floridana*, all the three functions i.e. leaf development, ICS and male fertility are carried out by a unique gene *MPF2* (He and Saedler, 2005). Hence, isolation of "evolutionary" mutants deficient in any of the three functions is highly unlikely. However, Withania being a tetraploid, as evident from ploidy evaluation, may offer the possibilities to sort out these functions. For this purpose, *MPF2-like* genes from Withania and Tubocapsicum were isolated, sequenced and phylogenetically as well as functionally analyzed.

The phylogenetic analyses have been done in collaboration with Dr. Jinyong Hu, (MPIZ, Cologne, former group member) and Simone Riss and Dr. Chaoying He (group memebrs) cooperated in Southern hybridization.

3.2.1 Withaninae contain two types of *MPF1-like* and *MPF2-like* genes

cDNAs of *MPF1-like* and *MPF2-like* genes were isolated from Tubocapsicum and 13 Withania accessions representing 5 Withania species listed in Table 3.3. For each species and for each of the genes at least two different types of clones were isolated compatible with the tetraploid nature of the species. Since the two types are both closely related to either *MPF1* or *MPF2* they were termed *MPF1* (or *2*)-*like A* and *MPF1* (or *2*)-*like B*, respectively. In some species such as *W. aristata* (W001) and *W. coagulans* (W002), *W. sp.* (W010 and W011) and *W. somnifera* (W012) more than two sequences for each of the two genes could be found. All the genes encode a highly conserved MADS-box, a short I-region, a K-box and a C-terminal region (Fig. 3.7). Besides, multiple clones of pseudogenes in both categories were isolated having a premature stop codon (Table 3.3). Pseudogenes (*PWC202_b*, *PWF203_a*, *PWR205_b*, *PWS206_b*, *PWF213_b*) were found in various accessions after two copies of *MPF2-like* genes were isolated.

5'-RACE was performed to obtain the full-length cDNA sequences of *MPF2-like* genes. The longest *MPF2-like-A* cDNA comprises 1107 bp, whereas the equivalent *MPF2-like-B* sequence is 1228 bp long. Both encode predicted polypeptides of 254 and 249 amino acid residues, respectively (Suppl. Fig. 1). Sequence comparison indicated that coding regions shared 89% identity (85 single-nucleotide differences in the 762 bp coding region) at the nucleotide level, whereas, their protein products are surprisingly, only 83% identical, (45 differences in 254 amino acid residues). This comparatively low conservation at protein level might allude towards the different functions for these two groups of *MPF2-like* genes.

Identifier	Species	MPF1-like		MPF2-like	
		Gene code	# clones	Gene code	# clones
T 001	*Tubocapsicum anomalum*	TAA101	7	TAB201_1	8
		TAB101	9	TAB201_2	14
W001	*Withania aristata*	WAA101_1	3	WAA201_1	3
		WAA101_2	10	WAA201_2	3
		WAB101	1	WAB201_1	3
				WAB201_2	8
W002	*Withania coagulans*	WCA102_1	3	WCA202_1	2
		WCA102_2	2	WCA202_2	15
		WCA102_3	10	WCB202	3
				PWC202_b	5
W003	*Withania frutescens*	WFA103_1	1	WFA203	4
		WFA103_2	2	WFB203_1	8
		WFA103_3	12	WFB203_2	2
		WFB103	2	PWF203_a	3
W004	*Withania riebeckii*	WRA104_1	1	WRA204_1	2
		WRA104_2	10	WRA204_2	4
				WRB204_1	9
				WRB204_2	2
W005	*Withania riebeckii*	WRA105_1	1	WRA205_1	3
		WRA105_2	10	WRA205_2	4
				WRB205	6
				PWR205_b	2
W006	*Withania somnifera*	WSA106_1	1	WSA206_1	3
		WSA106_2	10	WSA206_2	4
				WSB206	6
				PWS206_b	2
W007	*Withania somnifera*	WSA107	9	WSA207_1	3
		WSB107	3	WSA207_2	4
				WSB207_1	1
				WSB207_2	10
W008	*Withania somnifera*	WSA108_1	1	WSA208_1	1
		WSA108_2	2	WSA208_2	4
		WSA108_3	2	WSB208_1	4
		WSA108_4	2	WSB208_2	3
W009	*Withania sp.*	WPA109_1	2	WPA209	2
		WPA109_2	2	WPB209_1	3
		WPA109_3	2	WPB209_2	3
				WPB209_3	4
W010	*Withania sp.*	WPA110_1	1	WPA210	6
		WPA110_2	9	WPB210_1	7
		WPB110_1	1	WPB210_2	4
		WPB110_2	2		
W011	*Withania sp.*	WPA111	12	WPA211_1	3
		WPB111	1	WPA211_2	2
				WPB211_1	8
				WPB211_2	2
W012	*Withania somnifera*	WSA112_1	2	WSA212_1	3
		WSA112_2	8	WSA212_2	4
				WSB212_1	8
				WSB212_2	2
W013	*Withania frutescens*	WFA113_1	2	WFA213_1	3
		WFA113_2	1	WFA213_2	2
		WFA113_3	9	WFA213_3	2
		WFB11	2	WFB213	5
				PWF213_b	3

Table 3.3: List of taxa and genes used in the phylogenetic analysis. The source of species is given in Table 3.1 and Suppl. Table 1 (Hu and Saedler, 2007). *MPF1-like* and *MPF2-like* clones are abbreviated as 1 and 2, respectively. P in *MPF2-like* clones represents products from pseudogenes, featuring at least 1 premature stop codon in their sequence.

3.2.2 Tubocapsicum has lost *MPF2-like-A* genes

Although more than 20 *MPF2-like* cDNA clones were isolated and sequenced from Tubocapsicum, only *MPF2-like-B* genes could be identified indicating that *MPF2-like-A* genes might be missing in the genome (Table 3.3). Southern blotting experiment verified this assumption (Fig. 3.5). In Withania species one hybridization signal was detected if probed with a Withania *MPF2-like-A* gene cDNA, while no such signal was obtained with Tubocapsicum genomic DNA.

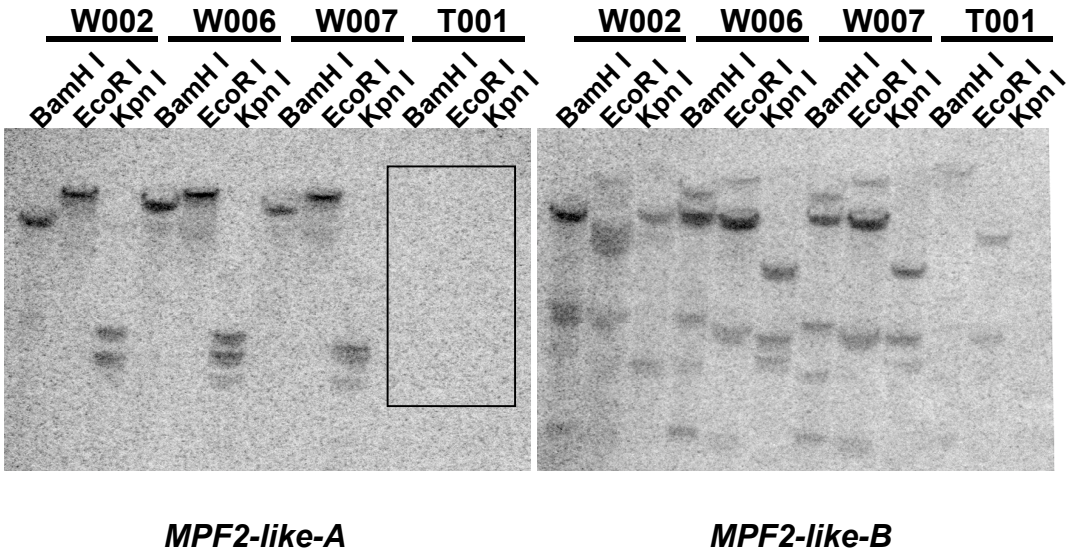


Figure 3.5: Southern blot analysis of *MPF2*-like genes in Withaninae

The genomic DNAs of W002, W006, W007 and T001 were digested with restriction enzymes indicated above and were hybridized with specific gene probes from the C-terminal regions of *MPF2-like-A* and *MPF2-like-B* from W006, respectively. Box highlights the missing signals in Tubocapsicum probed with *MPF2-like-A* genes.

This explicitly indicated a loss of *MPF2-like-A* gene in Tubocapsicum (Fig. 3.5). For *MPF2-like-B* gene the scenario was altered. At least one but often more signals were visible in Southern hybridization experiments. As expected for a polyploid species multiple alleles could be present in the genome. However, the intensity of *MPF2-like-B* signals in Tubocapsicum was weak. This could be explained in terms of the probe specificity. The probe for *MPF2-like-B* was designed in the highly specific region comprising C-terminal and 3-' UTR of *WSB206* gene. Consequently, weak signal were naturally predictable from the Tubocapsicum DNA.

After the isolation of different groups of *MPF1-like* and *MPF2-like* genes from the Withaninae, next step concerns the phylogenetic relationship of the species involved.

3.2.3 Phylogenetic tree reconstruction

Since *MPF2-like-A* is missing in Tubocapsicum only *MPF2-like-B* sequences were used in the following phylogenetic tree reconstructions. Concatenated DNA sequences of chloroplast markers *atpB* and *matK* and the nuclear cDNA sequences of *MPF1-like-B* and *MPF2-like-B* genes were used from *W. aristata* (W001), *W. coagulans* (W002), *W. frutescens* (W013), *W. riebeckii* (W004), *W. somnifera* (W006) and *T. anomalum* (T001) with *Iochroma australe* as out-group. A neighbor-joining tree is shown in figure 3.6.

High bootstrap values undermine the branching that places Tubocapsicum in the basal position to the Withania species. Both the Canarian species i.e. *W. frutescens* and *W. aristata* are sister to each other. These two plants species are also self-incompatible. Calyx morphologies in flowering and fruiting stages have already been explained in previous sections (see section 3.1). While 3 of the 5 Withania species share an open ICS, in *W. somnifera* and in *W. coagulans* a classical closed ICS is observed.

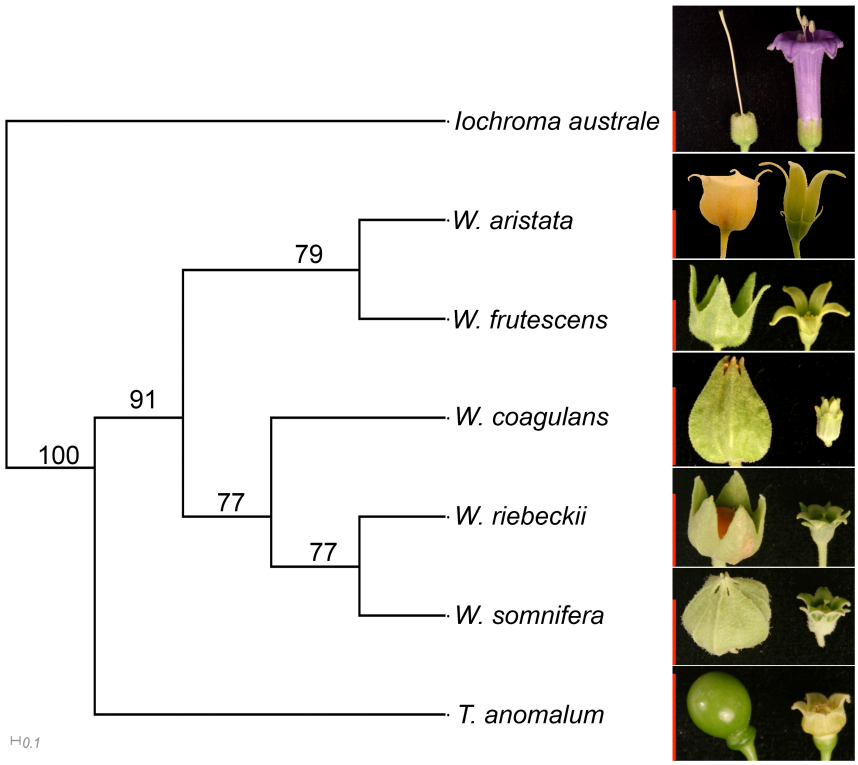


Figure 3.6: Phylogenetics analysis of Withaninae. A neighbor-joining tree of 5 Withania and 1 Tubocapsicum species, based on concatenated sequences of *matK*, *atpB*, *MPF1-like-B* and *MPF2-like-B* sequences, was established using *Iochroma australe* as out-group. Details are described in Materials and Methods. The morphologies of calyces in flower and fruit stages are included. The scale bar is given as vertical red line with fruiting calyx. Bar = 1cm. The horizontal bar indicates the substitution rate. Bootstrap values are given above the branches.

The two types of *MPF2-like* orthologs, the *MPF2-like-A* and *MPF2-like-B* genes can easily be distinguished in their protein alignments.

3.2.4 MPF2-like Protein structure have diverged into MPF2-like-A and MPF2-like-B proteins

Figure 3.7 depicts an alignment of MPF2 of *Physalis floridana*, its orthologs STMADS16 of *Solanum tuberosum*, MPF2-like-A and MPF2-like-B of *Withania somnifera* and MPF2-like-B of *Tubocapsicum anomalum*. Two cDNA clones of Withania and Tubocapsicum each are shown, revealing the heterozygosity in each gene of these plants.

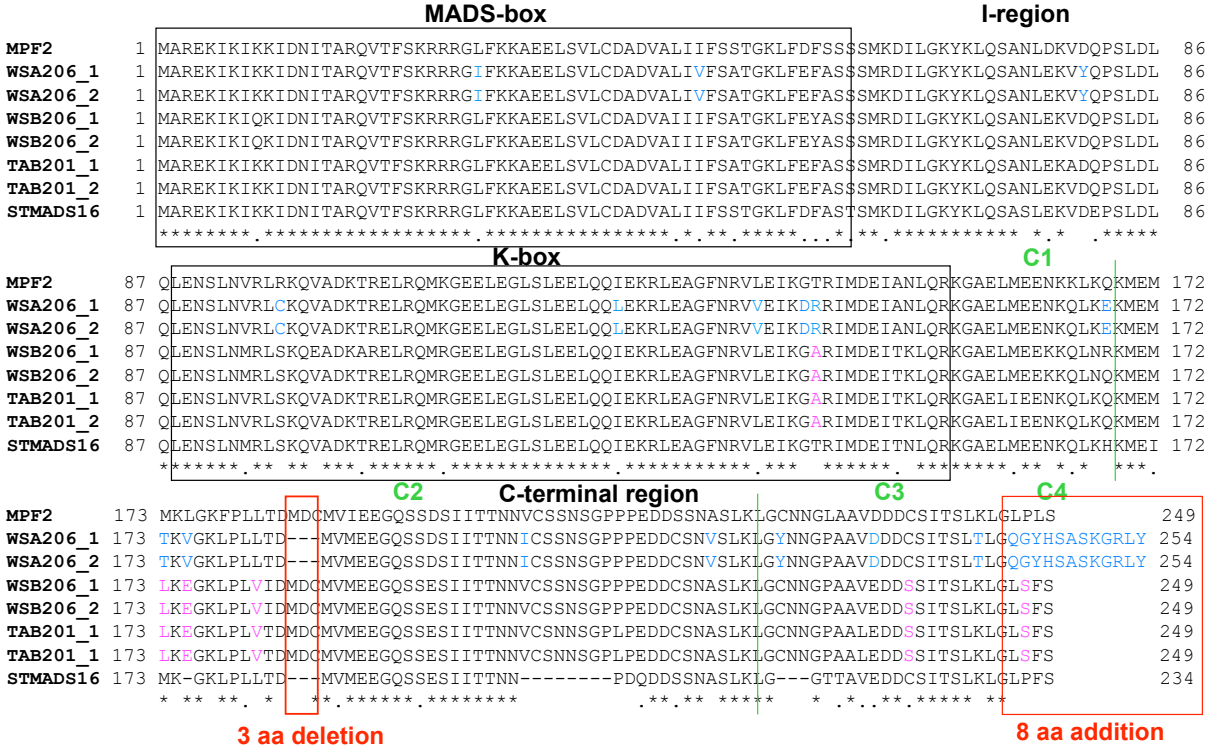


Figure 3.7: Structural comparison of MPF2-like sequences. Deduced amino acid sequence comparisons of MPF2-like proteins. The four domains of the MADS-box proteins are indicated above the sequences. The aa in blue indicate the WSA206 specialized amino acid positions; in blue bold show the positively selected amino acids position detected with PAML (phylogenetic analysis by maximum likelihood) analysis; aa in pink mark the WSB206 specialized amino acids which are different from STMADS16; green lines mark the exon borders of C-terminal region; red boxes mark the two significant structural changes, the 3aa deletion and 8aa addition in WSA206.

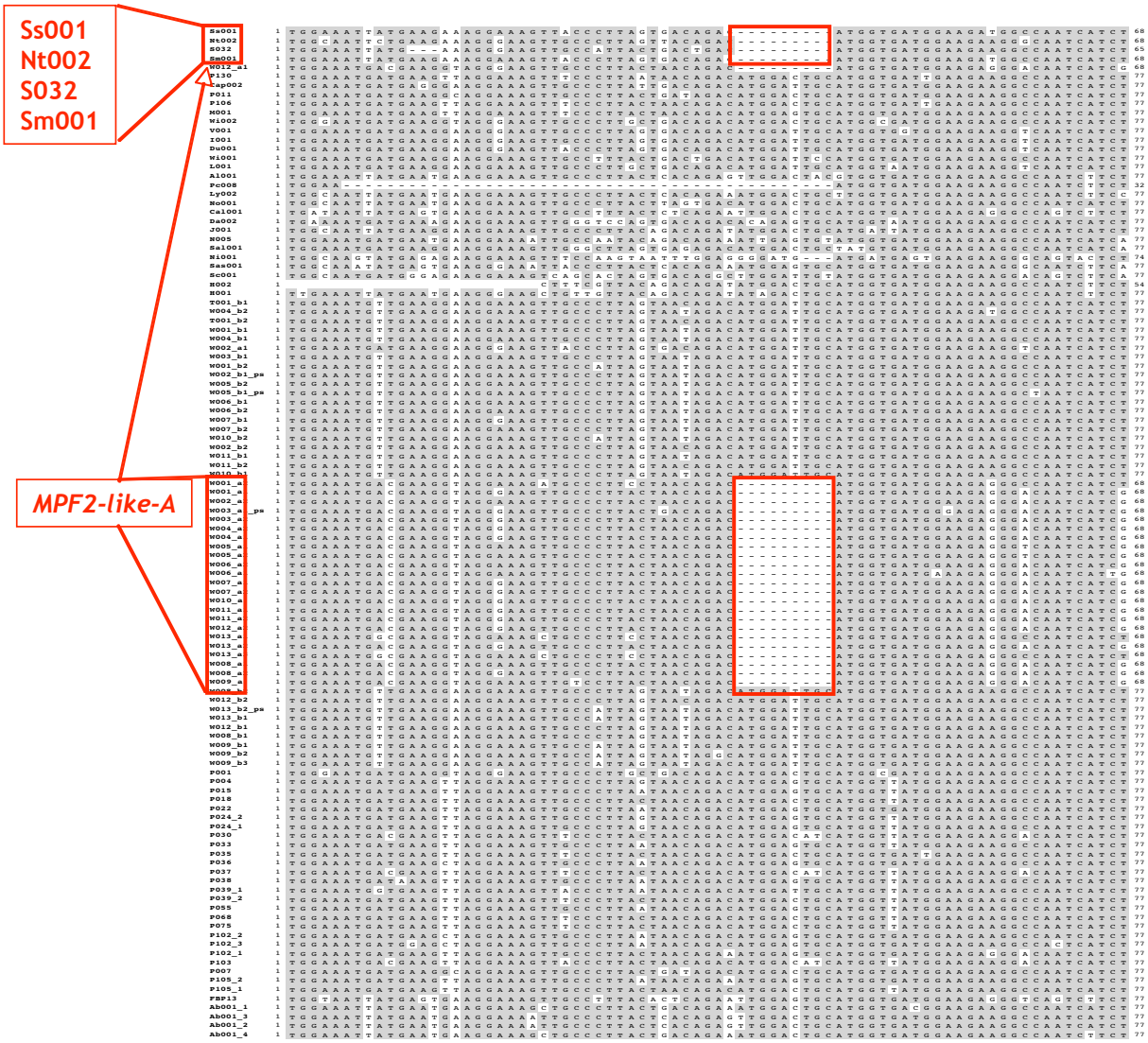


Figure 3.8: Data matrix with ClustalW alignments of C-terminal region of *MPF2-like* genes. Red box indicates the missing 9 nucleotides (3aa) in *MPF2-like-A* genes as well as in the Ss001 (*Solanum sisymbriifolium*), Nt002 (*Normania triphylla*), S032 (*Solanum tuberosum*) and Sm001 (*Solanum macrocarpon*).

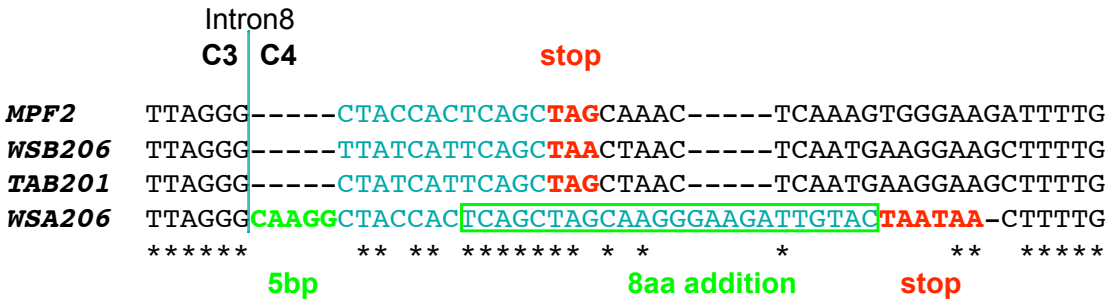


Figure 3.9: The MPF2-like-A specific 8 aa addition could be due to frameshift. Nucleotide alignments at 3rd and 4th exons of C-terminal region of *MPF2-like* and *MPF2* genes. C3 and C4 are the exons and pink vertical line demarcates the C3 exon from intron 8. Red letters indicate the stop codons. 5bp insertion (CAAGG) is shown in green letters. Green box encloses the 24 nucleotides or the 8aa addition in *WSA206*.

Even though there are amino acid exchanges in the MADS-box, I-region and K-box, the two most prominent differences compared to MPF2 are restricted to the C-terminal region. A 3 amino acid deletion has occurred in MPF2-like-A proteins in the second exon (indicated by the red box). However, this deletion is not specific to MPF2-like-A proteins. It is already found in STMADS16 and other Solanum MPF2-like proteins (Fig. 3.8) including Ss001 (*Solanum sisymbriifolium*), Nt002 (*Normania triphylla*) and Sm001 (*Solanum macrocarpon*). The other striking change occurred at the end of C-terminal region, which is the 4th exon of C-terminal region. At the carboxy terminal end following a 4 amino acid sequence diversity an extension of 8 amino acids is observed (all are included by the red box). Therefore, this change can be interpreted as an 8 aa addition. This 8 aa addition is specific to MPF2-like-A protein and is the hallmark of this protein. Furthermore, the nucleic acid sequences of the 3rd and 4th exons as well as the 3' un-translated region (3' UTR) were compared. Clearly, a 5bp long insertion (CAAGG) has resulted in the reading frameshift, thus altering the position of the stop codon (Fig. 3.8). Besides these two dramatic changes, other discernible alterations could be found in the 3' UTR region.

Though noticeable changes in MPF1-like proteins could also be detected in the fragment obtained, these changes were not as striking as for the MPF2-like proteins (data not shown).

3.2.5 Gene structures of *MPF2-like* genes are largely maintained

In order to define the diversification of exon-intron structures of *MPF2-like* genes, genomic sequences of two representatives of the Withaninae including *T. anomalum* (T001) and *W. somnifera* (W006) with *MPF2* from *P. floridana* as reference were compared (Fig. 3.10). All the *MPF2-like* genomic loci tested comprise one exon of MADS-box, one exon of I-region, 3 exons of K-box and 4 exons of C-terminal region with 8 introns separating them. Fig. 3.10 demonstrates that gene structures of the 4 groups of genes have been basically maintained with some variability in the lengths of exons and introns. For instance, the length of exons shows only the changes as described above concerning the 3aa deletion and an 8 aa addition in *MPF2-like-A* genes. However, the length of introns varies considerably. The first introns of *WSA206*, *MPF2*, *WSB206* and *TAB201* are 2565, 2158, 1820 and 3617 base pair long, respectively. Remarkably, the second intron of *WSA206* (1043bp) is much longer than that of *WSB206* (422bp) and *TAB201* (422bp) but similar to *MPF2* (832bp). *WSA206* has a much shorter 3rd intron (922bp) than all the other introns (around 1.5kb). Other introns also changed a little.

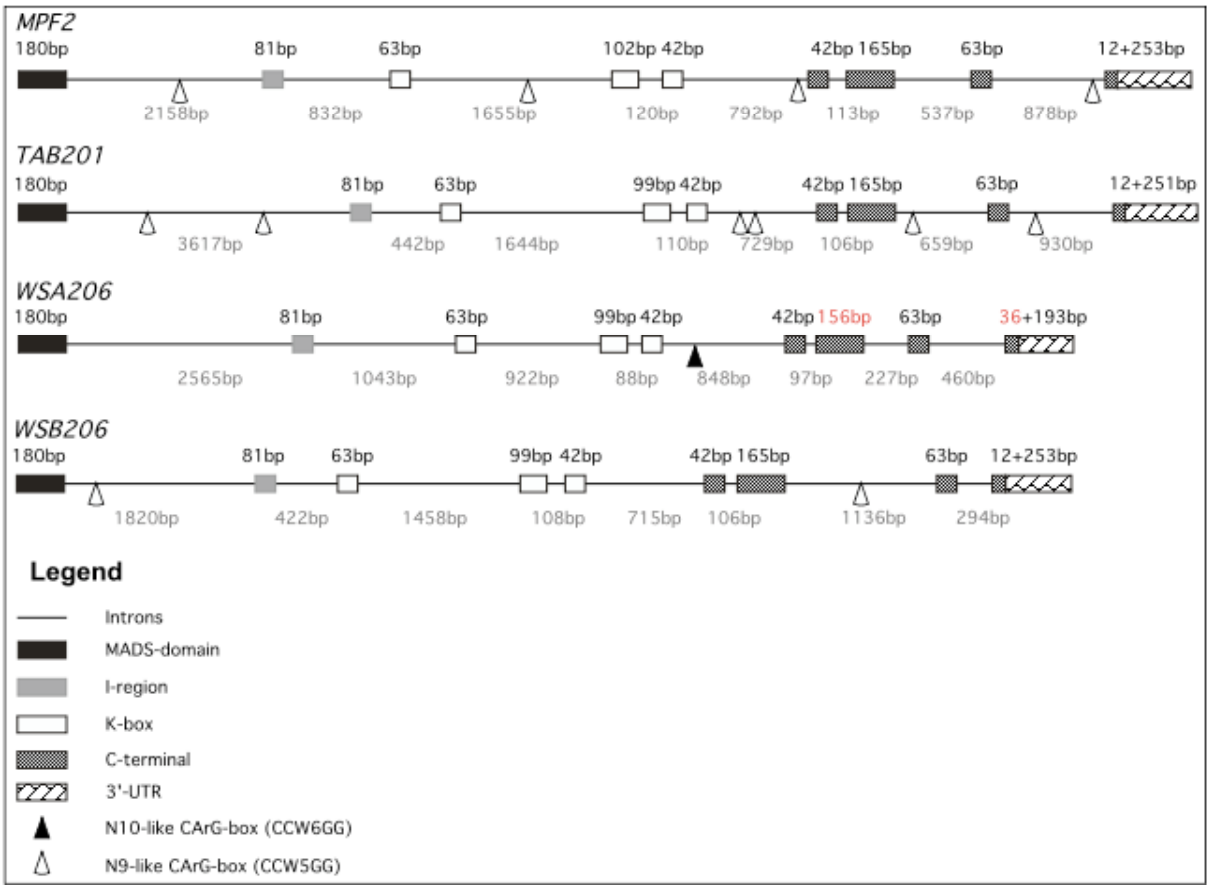


Figure 3.10: Exon-intron structures of *MPF2-like* genes

Numbers above the lines give exon length while numbers below the line give the intron length. The red number 156bp in *WSA206* gene indicates the 9bp (3 amino acids) deletion in the C-terminal region and the red number 36 symbolizes a 24bp (8 amino acids) addition at the end of C-terminal region of this gene. Legend is also given.

In essence, at the carboxy terminal end following a 3 amino acid deletion an 8 amino acids extension is observed. These two structural features are signatures of *MPF2-like-A*-genes of Withania species. In principle both *MPF2-like* genes could be responsible for ICS formation in Withania. However, non-existence of 3aa deletion and 8aa extension suggests that *MPF2-like-B* seems to be more related to *MPF2*, which in turn also might suggest functional equivalence.

Phylogenetic analyses of these two groups of genes can provide further insights into the differentiation of these two groups of genes in the Withaninae.

3.2.6 *MPF2-like* genes have duplicated probably due to genome duplication

To reveal the evolutionary history of both *MPF1-like* and *MPF2-like* genes in Solanaceae, neighbor-net network analysis was employed. A broader sample of *MPF2-like* cDNA sequences from various Solanaceae (Hu and Saedler, 2007) plus the 10 accessions of 5 Withania species, 3 unidentified Withania species and 1 accession of *Tubocapsicum anomalum* (Table 3.3) were subjected to neighbor-net network analysis. Two clusters of *MPF2-like* genes are observed in Withania, containing either *MPF2-like-A* or *MPF2-like-B* genes, respectively, and are boxed in blue and red in Fig. 3.11.

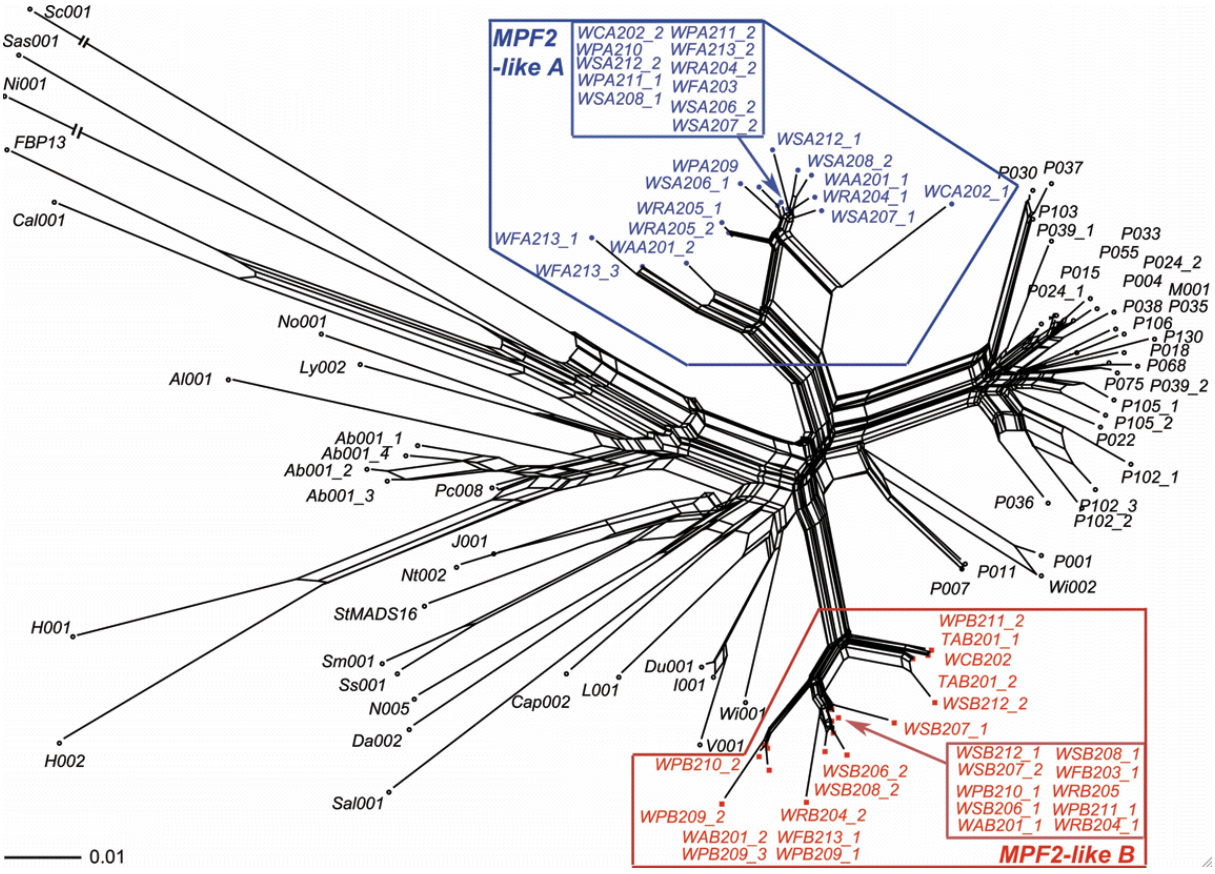


Figure 3.11: Phylogenetic networks of *MPF2-like* genes of Solanaceae. Genes are given using species identifiers (Hu and Saedler, 2007), except for the Withaninae genes (Table 3.3). The *MPF2-like-A* genes are boxed in blue and the *MPF2-like-B* genes in red box. The bar represents 1% mutation rate.

Each of the gene clusters features 3 sub-groups. These might reflect gene duplications or very diverse alleles present in the corresponding species. For example *Withania aristata* contains 2 well-separated *MPF2-like-A* genes (*WAA201-1* and *WAA201-2*), *W. frutescens* features *WFA213-1/3* and *WFA213-2* and *W. coagulans* contains *WCA201-1* and *WCA201-2*. Multiple clones from other species tend to cluster and hence their diversity might represent different alleles. In the *MPF2-like-B* gene cluster the situation is similar. Duplications might have occurred in *W. aristata* (W001) and *W. coagulans* (W002), but also in the *W. sp.* (W010 and W011) and in *W. somnifera* accession W012, while in the other species again clustering is observed. All *MPF2-like-A* of the four *W. somnifera* accessions cluster as do 7 out of 8 *MPF2-like-B* genes suggesting that the diversity observed within this species very likely reflects different alleles.

A similar phylogenetic network was observed with *MPF1-like* cDNA sequences (Table 3.3). Again 2 well-separated clusters, *MPF1-like-A* and *MPF1-like-B*, were observed in the Withaninae (Fig. 3.12), however, unlike in the case of *MPF2-like* genes, Tubocapsicum also contains both *MPF1-like-A* and *MPF1-like-B* genes.

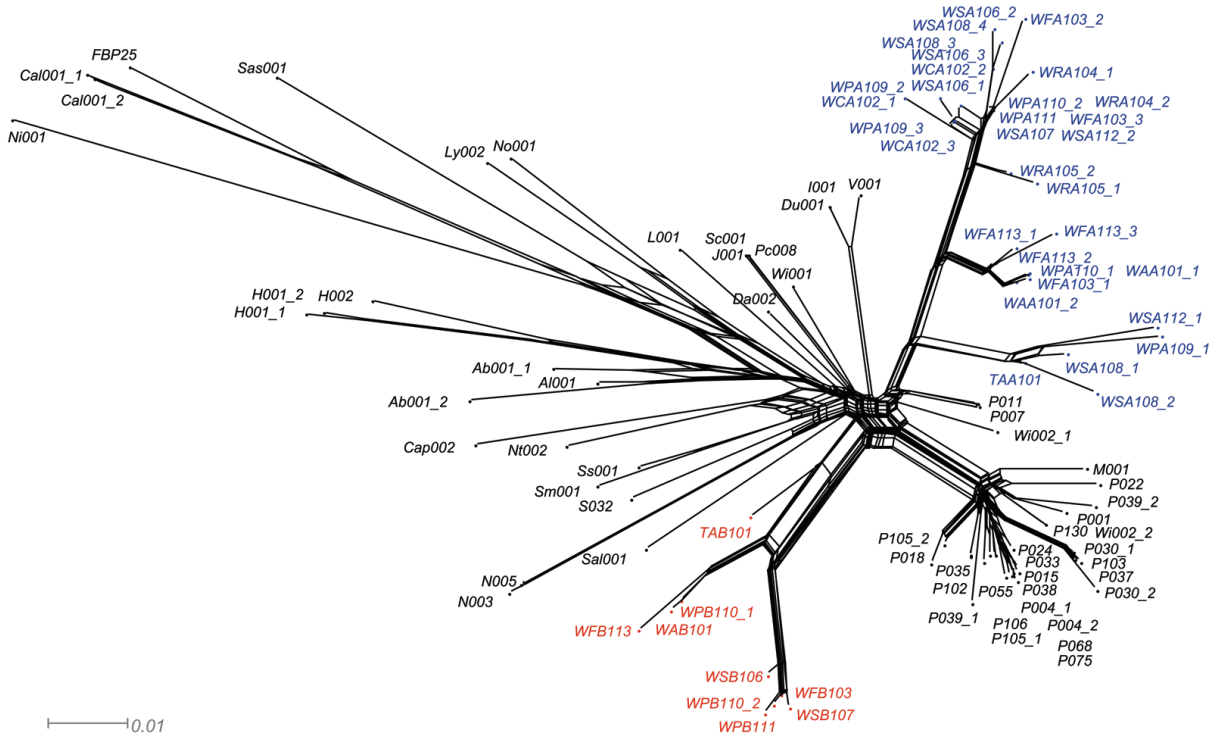


Figure 3.12: Phylogenetic net network of *MPF1-like* genes of Solanaceae. Phylogenetic net network of *MPF1-like* genes of Solanaceae is shown. Genes are given using species identifiers (Hu and Saedler, 2007), except for the Withaninae genes (Table 3.3). The *MPF1-like-A* genes are in blue and the *MPF1-like-B* genes in red. The bar represents 1% mutation rate. Note both the *MPF1-like-A* and *MPF1-like-B* genes in Tubocapsicum.

This clustering clearly suggested the duplication or diversification of Withania *MPF2-like* genes. This duplication was very likely due to the genome duplication, since *Withania* and Tubocapsicum were proved to be tetraploids.

As Tubocapsicum features no ICS and has lost the *MPF2-like-A* gene, it can be hypothesized that *MPF2-like-A*, given the peculiarities of its amino acid sequence in its C-terminal region, seems to be responsible for ICS formation in Withania. If so, this might indicate that two proteins were under different selection pressures.

3.2.7 MPF2-like-A proteins are positively selected during evolution

To examine whether MPF2-like proteins are subjected to positive Darwinian selection, a phylogeny based statistical test was performed. Darwinian selection has been studied previously in several cases (Pogson and Mesa, 2004; Spillane et al., 2007, Des Marais and Rausher, 2008). Therefore, directional evolution of MPF2-like-A and MPF2-like-B proteins was tested in the Withaninae and MPF2-like proteins in Physalis. Since the MPF2-like proteins of Iochroma, Dunalia, Vassobia, *Solanum tuberosum* and *Solanum macrocarpon* are basal to both Withaninae and the Physalis proteins, they were taken as the pro-orthologs (Bielawski and Yang, 2004) and the calculations were done according to Yang and Bielawski (2000).

Branch models were used to check whether different ω-ratios along different lineages in the phylogeny could be detected and hence might suggest different selection pressures on lineage-specific genes after duplication. The M0 and free-ratio model with 1, 2 and n (the total number of lineages) ω-ratios was used. A two-ratio model that defined a single ω-ratio for the branches leading to MPF2-like-A proteins (suggesting diversifying function), and a second ω-ratio for all the other lineages including the pro-orthologues as well (assuming functional conservation), provided a significantly better fit for the data than the one-ratio model with a single ω-ratio for the complete phylogeny ($P=0.003$). Another two-ratio model that assumed only *MPF2-like-B* genes with functional diversification was not better than the one-ratio model ($P>0.1$).

The results suggest that the branch leading to the MPF2-like-A proteins in Withaninae and the MPF2-like proteins in Physalis seems to have been under positive diversifying selection (highlighted in bold red number, Fig. 3.13), while all other branches including the one leading to MPF2-like-B proteins are under purifying selection.

The positive selection of Physalis MPF2-like and Withania MPF2-like-A proteins correlates with ICS formation in both genera.

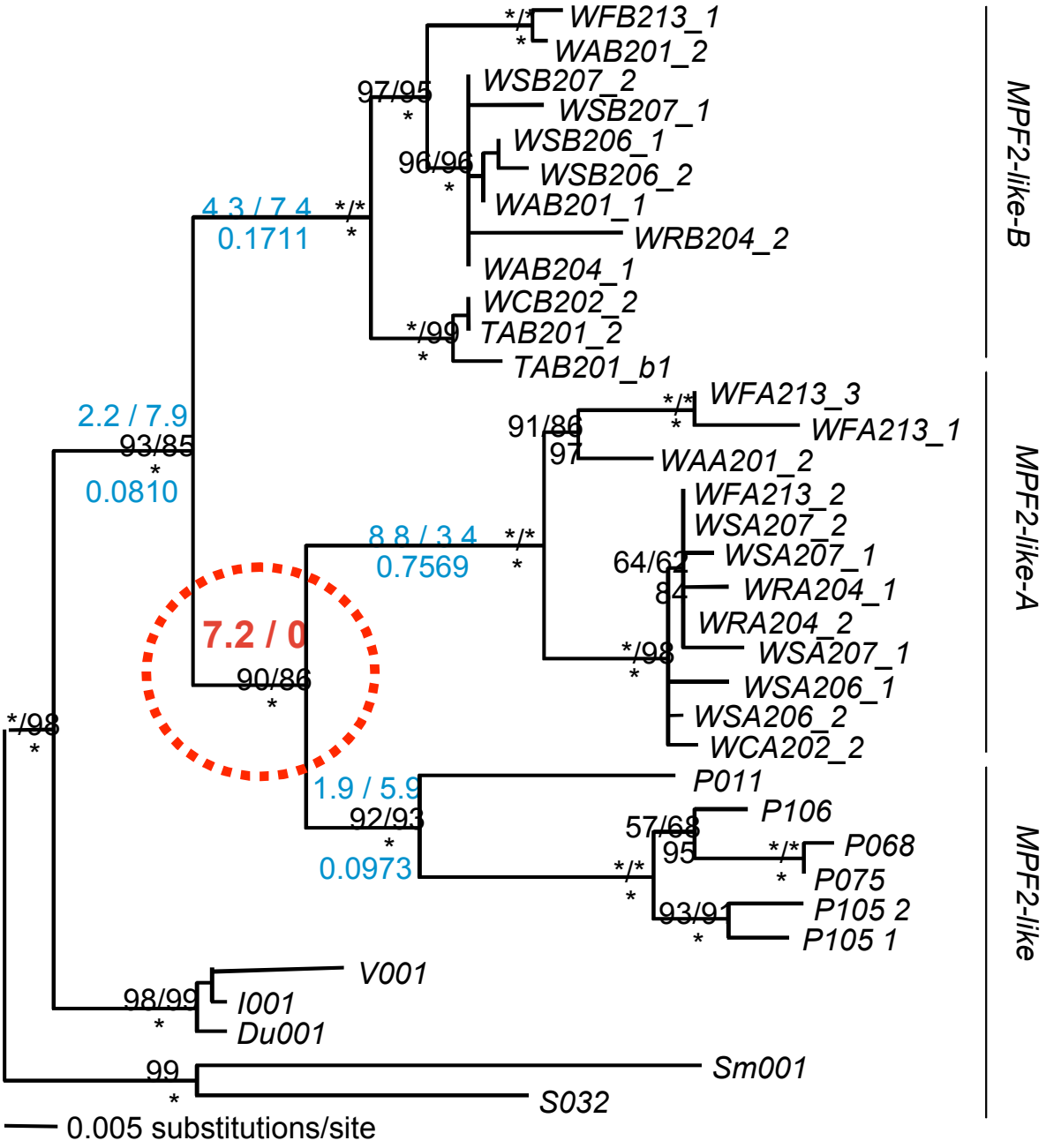


Figure 3.13: Darwinian selection of MPF2-like proteins in Physalinae

Numbers above the major branches of the unrooted ML tree are averages of non-synonymous and synonymous changes, while ω = dN/dS values are given below branches. ω>1, positive selection; ω=1, no selection; ω<1, purifying selection. Numbers highlighted bold in dotted circle indicate the branch underlying positive Darwinian selection. Bootstrap values for MP, ML and Bayesian posterior probability values of the major branches were 85%. * denotes the Bayesian posterior probability value.

3.2.8 Identification of amino acid positions under positive selection

Because MADS-domain proteins have 4 important domains showing different capabilities in the protein function, it was tested whether positive selection has occurred to certain domains and/or at certain amino acid positions along the MPF2-like proteins. Phylogenetic analysis by maximum likelihood (PAML) suggested that 5 amino acid positions were positively selected (see Fig. 3.7 protein sequence alignment). These included the #28 (I) in the MADS-box, #126 (L), #138 (V) and #143 (R) in the K-box and #234 (D) of C-terminal region (see bold letters in Fig. 3.7). This statistical analysis can provide the evidence for positive selection but is unable to shed light on its mechanism. However, pinpointing amino acid sites likely to be involved in positive selection hint clues for laboratory investigations.

These results suggested that the positive selection was mainly restricted to K-box of the MPF2-like proteins besides some selection in both MADS-box and C-terminal region. In addition, the 3aa deletion and 8aa addition might very likely also contribute to the functional divergence of the duplicated genes.

3.2.9 Diverged Darwinian selection of MPF1-like proteins

No significant differences among MPF1-like proteins post-duplication via comparing the one-ratio model with a free-ratio model (P=0.3827) were found. This indicated that there was no reason to reject the null hypothesis that MPF1-like proteins have evolved at constant rates along lineages. However, a highly diverged pattern of selection pressure on MPF1-like proteins through lineage-specific analyses of non-synonymous versus synonymous change via free-ratio model for MPF1-like proteins was observed (Fig. 3.14). Contrary to MPF2-like proteins, where Darwinian positive selection was mainly acting on the lineage leading to Withania MPF2-like-A proteins and Physalis MPF2-like, in this analysis the Darwinian positive selection was detected in various deep lineages, for examples, the branch leading to both the MPF1-like-A and MPF1-like-B proteins. The ω ratio for each branch changed drastically. Contrary to MPF2-like proteins, the lineage consisting of the Iochrominae proteins of *MPF1-like* have variable ω ratio after separation from their common ancestor. Therefore, MPF1-like proteins have apparently evolved in a very different way compared to MPF2-like following duplication.

Since the MPF2-like-A proteins were positively selected during evolution of the Withaninae, the expression of this gene also might be different compared to that of MPF2-like-B.

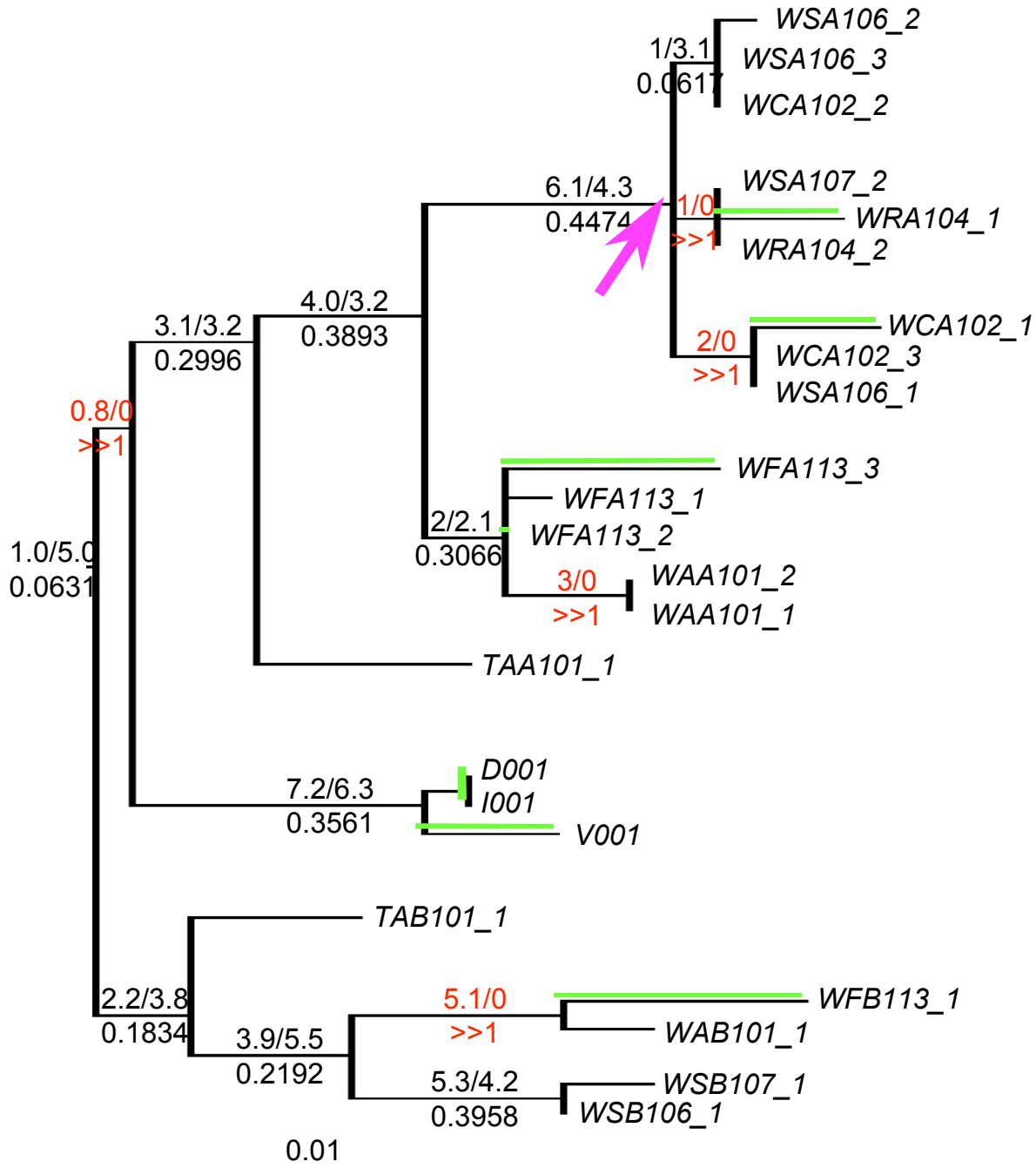


Figure 3.14: Diverged Darwinian selection of MPF1-like proteins

Above branches give the non-synonymous/synonymous substitutions; below branches give the ω (=dN/dS); in red indicates ω far above 1. ω>1, positive selection; ω=1, no selection; ω<1, purifying selection. Terminal branches in green show genes with free or positive mutations. The pink arrow indicates the cluster where all the pseudogenes are located.

3.2.10 *MPF2-like-A* and *MPF2-like-B* genes have different expression patterns

To analyze the functional divergence of *MPF2-like* genes after duplication, the temporal expression patterns of *MPF2-like* gene transcripts were determined by RT-PCR analysis (Fig. 3.15). All the *MPF2-like* genes were strongly expressed in leaf tissues irrespective of the developmental stage. The transcript signals of *Dunalia* and *Iochroma* were not detectable in flower tissues at different developmental stages while the expression of the *Vassobia* gene was found to be weak in the young flower buds and later disappeared in more developed flowers. These expression patterns are comparable to previously published reports by Hu and Saedler (2007).

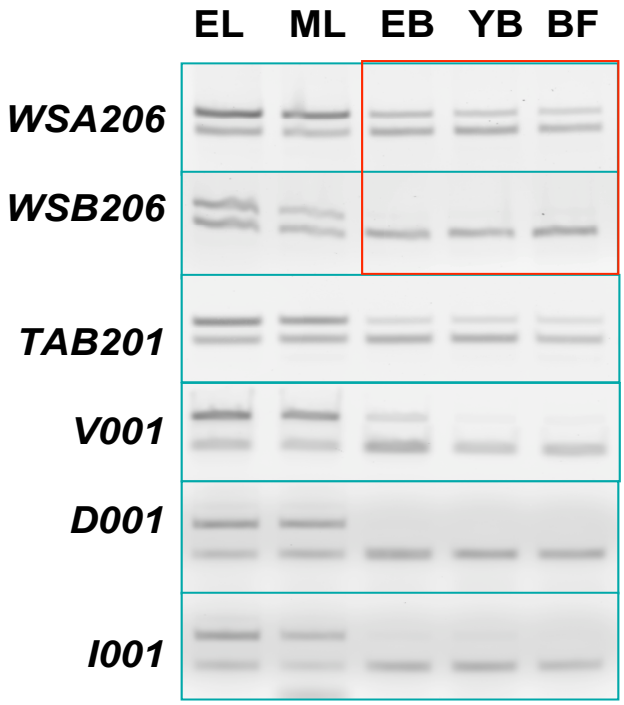


Figure 3.15: RT-PCR analysis of *MPF2-like* genes

RT-PCR of the genes is indicated in various tissues of the respective plants: EL, newly developed leaves (<1cm); ML, mature leaves; EB, early flower buds prior to egression; YB, young flower buds, pre-anthesis; BF, blooming flowers, anthesis. The upper band shows the expression of the *MPF2-like* gene, the lower band serves as control and represents 18S RNA. Red box highlights the specific expression of *WSA206* in flower and no expression of *WSB206* in the flower.

It is noteworthy that *WSA206* genes were also universally expressed in the flower tissues at different developmental stages. But strikingly, *WSB206* genes expression was barely detectable in flower at any stage. However, increasing the PCR cycles to 35 very weak signals of *WSB206* transcripts could be detected in flower. Contrary to Withania *MPF2-like-B* gene expression, Tubocapsicum *TAB201* was universally expressed both in both the leaf and flower tissues. Since no signal for *TAA201* gene could be detected in Southern blotting experiment it was not surprising that no *TAA201* transcript could be detected unambiguously. These results

revealed that the expression of *WSB206* is confined to leaves only while *WSA206* is strongly expressed in both the leaf and flower tissues.

The comparison of cDNA sequences, amino acid sequences, gene structures and expression patterns demonstrated that there are two different groups of genes in *Withania*. Given the restricted expression of *WSB206* compared to *WSA206* and the structural difference of their proteins one could strongly suggest that they also differ in functionallity.

3.2.11 Functional differentiation of the MPF2-like-A and MPF2-like-B proteins

Previously it was shown that over-expression of *MPF2* and *STMADS16* genes in Arabidopsis leads to increased secondary sepal growth only in the presence of external cytokinins application, while MPF2-like proteins do not require the external hormones applications to generate secondary sepals growth (He and Saedler, 2007). To elucidate the functional differentiation of MPF2-like proteins, WSA206 and WSB206 from *W. somnifera* and TAB201 from *T. anomalum,* therefore, were over-expressed in *Arabidopsis thaliana* using the CaMV promoter.

Plant transformation efficiency was very high in all the constructs including empty vector construct (wild type). The presence of transgene was confirmed by RT-PCR analysis on cDNA made after extracting RNA from transgenic Arabidopsis plant seedlings. The presence of 680 bp PCR product in Fig. 3.16 in all the transgenic plants lines for their corresponding genes (*WSA206*, *WSB206*, *TAB201*) verified their transgenic nature. At flowering stage, transgenic plants were compared to each other and with wild type plants to observe the differences in phenotypes pertaining to inflorescence, flower and fruit development (Fig. 3.17A-C).

3.2.12 *MPF2-like-A* over-expressor mimicked the *35S:AGL24* phenotype

Very obvious phenotypic changes were manifested by the plants transformed with *WSA206* genes in both the vector backbones i.e. PXCS-mYFP and pBAR-A. The most spectacular phenotypic alteration was the enormous growth of secondary sepals in flower (Fig. 3.17A). Transgenic plants obtained from *35S:WSA206* transformation tagged with and/or without YFP showed increase in sepal growth. However, the enlarged sepals phenotypic efficiency was higher (24%) in plants, in case of YFP tagged *WSA206* gene, in comparison with 11% without YFP tag. Along with increase in size, the sepals became persistent indicating loss of abscission at flower abscission zone or receptacle. Loss of abscission might be in turn necessary to prevent the sepals from dropping off and eventually facilitating the sepal growth.

Moreover, the sepals became more fleshy and curved inwards at the tip to form a lantern like structure (Fig. 3.17A,L). Branched trichomes gave the sepals a typical leaf like appearance. These phenotypes mimicked the *35S:AGL24* over-expressor in Arabidopsis (He and Saedler, 2007). But it is a bit different from *AGL24* over-expressor in not having a secondary flower in the axils of stamens (Fig. 3.17I,J). The enormous growth of sepals seems to be pressing the organs in the inner whorls, the petals particularly (Fig. 3.17E).

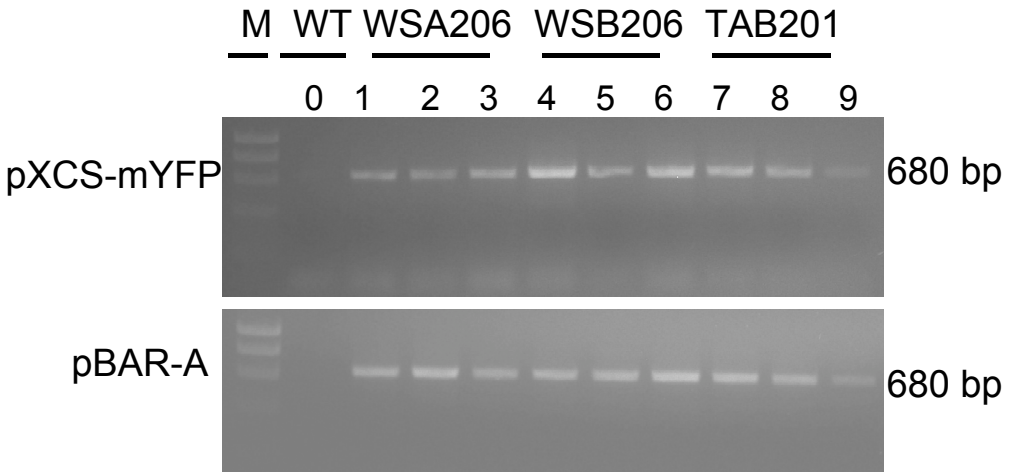


Figure 3.16: Confirmation of *MPF2-like* transgenes in Arabidopsis seedlings by RT-PCR analysis

The upper panel shows RT-PCR analysis of transgenic plants seedlings generated by PXCS-mYFP vector backbone. The lower panel represents RT-PCR for pBAR-A vector backbone transgenic plants. Numbers (1-9) indicate transgenic lines used (in both the vector backbones) for RT-PCR analysis. The 0 symbolizes the wild type plants. Marker is denoted by M and three genes *WSA206*, *WSB206* and *TAB201* are shown above the corresponding transgenic lines.

In addition to increased sepal growth, other phenotypic alterations in inflorescence and fruit shape were exhibited by *35S:WSA206* Arabidopsis transformants. Among these an interesting phenotype was the tremendous increase in the length of pedicel bearing the inflorescence (Table 3.4). A maximum length of 21 mm and minimum of 16 mm is recorded in transgenic plants transformed with *WSA206* gene. On the other hand *35S:WSB206* shows only 11.06 mm, which is the average length close to pedicels length (10.01 mm) of wild type plants. In the case of Tubocapsicum gene transformants, an average of 12.00 mm increase in pedicel

Figure 3.17: Phenotypes exhibited by transgenic Arabidopsis plants transformed with *MPF2-like* genes

A) *35S:WSA206* early, middle and late stages of flower; B) *35S:WSB206* early, middle and late stages of flower; C) *35S:TAB201* early, middle and late stages of flower; D) Wild type early, middle and late stages of flower; E) *35S:WSA206* inflorescence; F) *35S:WSB206* inflorescence; G) *35S:TAB201* inflorescence; H) Wild type inflorescence; I) *35S:AGL24* inflorescence; J) *35S:AGL24* flower bearing a secondary flower in the axils of stamens; K) Silique comparison of *35S:WSB206* (upper) with *35S:WSA206* (lower); L) Silique comparison of *35S:WSA206* transformants. Bar = 5 mm.

length can be noticed which, is intermediate between *35S:WSA206* and *35S:WSB206* transformants.

A great degree of variability in silique size and shape was observed in plants transformed with *35S:WSA206* construct (Fig. 3.17L). The siliques became shorter but larger in width. It is noteworthy that in some of the lines the siliques became extremely small and did not contain any seed inside. These observations indicate that *WSA206* gene may partly contribute to the fertility characteristics in flower. Moreover, the surface of the siliques became wrinkled and grooves were clearly visible (Fig. 3.17K,L). The tip of the siliques was slightly bent to make a curve-like structure (Fig. 3.17I).

Contrary to *WSA206* transformants, no visible alterations in the phenotype could be observed in case of transgenic plants expressing *WSB206* or *TAB201* genes (Fig. 3.17B-D). These plants behaved like normal wild type plants. Although the sepal persistence could be observed in these plants but it could not bring about a noticeable increase in secondary sepal growth. Besides, no differences in the pedicel length, inflorescence character and fruit size and shape were noticed (Fig. 3.17B,C,D,F,G,H).

Leaves and sepals of transgenic plants transformed with *MPF2-like* genes were scanned with Laser Scanning Confocal Microscope (LSCM) to monitor the intensity and localization of YFP signals. Generally, the transgenic plants transformed with YFP gene driven under CaMV 35S promoter are envisioned to show strong YFP expression ubiquitously in all the tissues and cells. Although the sepals of *WSA206* transformants exhibited strong YFP signals in the nucleus (Fig. 3.18A,B,C) but signals in leaves were weak (Fig. 3.18G,H,I). Weak signals could also be detected in leaves and sepals of *35S:WSB206* also (Fig. 3.18D,E,F). But unexpectedly, the YFP signals from the plants transformed with empty vector constructs were also weakly detectable in the nucleus.

These results demonstrated that in *35S:WSA206* transgenic plants enlarged sepals are formed and these do not drop off during fruit development. These features clearly distinguish the *MPF2-like-A* and *MPF2-like-B* gene functions and therefore suggest sub- (neo)-functionalization of *WSA206* and *WSB206* during the evolution of the Withaninae. Moreover, enlarged sepal formation suggests that *MPF2-like-A* gene is involved in ICS formation.

As *MPF2-like* genes demonstrate diversity in expression patterns, a study of regulatory control elements can further substantiate the understanding of sub- (neo)-functionalization phenomenon in the Withaninae.

Lines	WT	WSA206	WSB206	TAB201
1	11	16	9	10
2	14	20	10	10
3	8	19	13	9
4	11	17	11	11
5	9	21	10	10
6	10	18	10	10
7	10	18	12	11
8	9	17	11	11
9	10	16	10	13
10	8	21	13	15
11	11	19	14	14
12	10	18	10	14
13	9	20	11	13
14	8	16	11	16
15	11	19	12	13
16	12	18	10	12
Average	**10.01**	**18.31**	**11.06**	**12.00**
St. Dev	**±1.61**	**±1.66**	**±1.39**	**±2.06**

Table 3.4: Measurements of pedicel lengths of transgenic Arabidopsis plants. Pedicel length is measured in millimetres. WSA206, WSB206, TAB201 and WT represent the transgenic plants lines transformed with these genes driven under 35S CaMV promoter along with empty vector as a wild type control. St. Dev. is the standard deviation of measured values of these transgenic lines. Note the average increase in pedicel length (18.31) in the case of *35S:WSA206* transgenic lines.

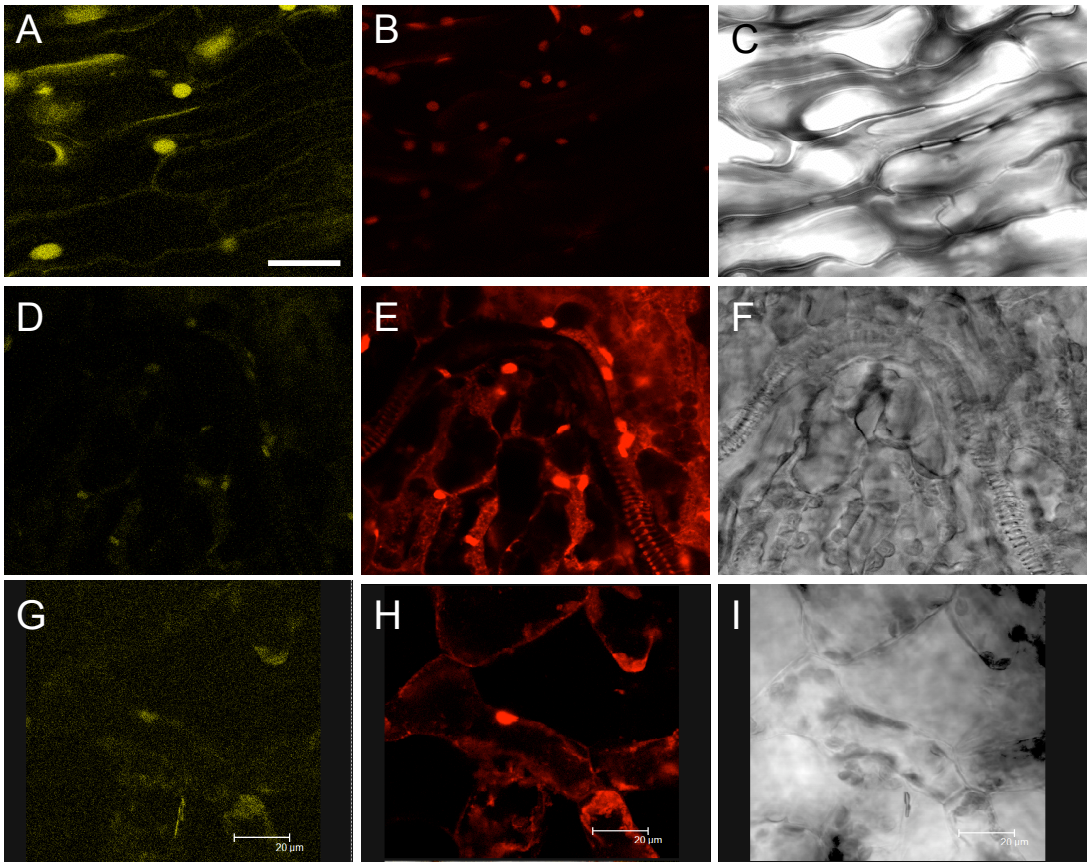


Figure 3.18: Intensity and localization of YFP signals in the transgenic Arabidopsis

A to C) *35S:WSA206* sepal in YFP, chlorophyll and white channels, respectively; D to F) *35S:WSB206* sepal in YFP, propidium iodide and white channels, respectively; G to I) *35S:WSA206* leaf in YFP, propidium iodide and white channels, respectively. Bar = 20 µm.

3.3 The *cis*-regulatory elements of *MPF2-like* genes

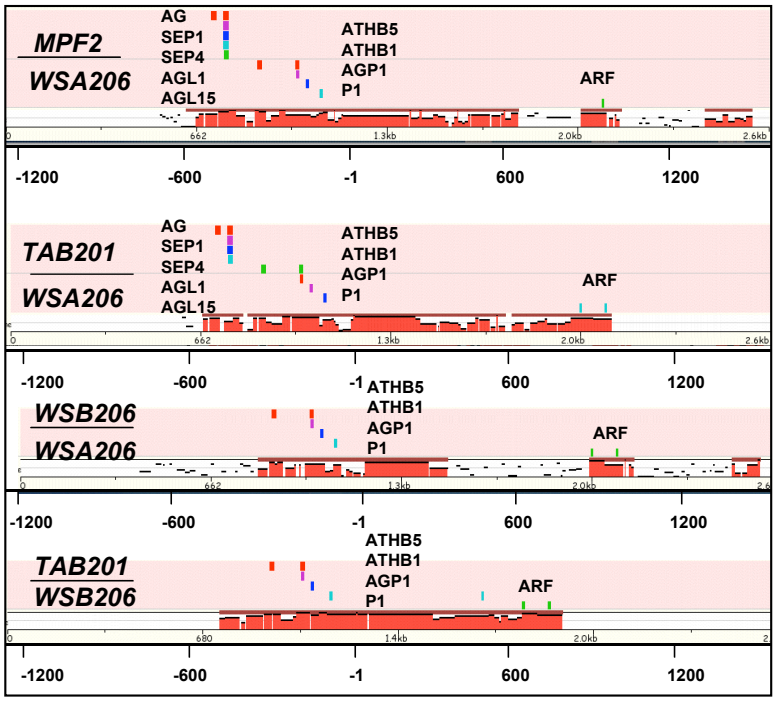
In the previous section RT-PCR analysis has revealed that the expression of *MPF2-like-A* genes are strongly detectable in both the leaves and flowers but the expression of *MPF2-like-B* is restricted to leaves only. This differential expression of *MPF2-like-A* and *MPF2-like-B* reflects differences in regulation of these genes. Changes in *cis*-acting or *trans*-acting factors are accounted for the appearence of morphological novelties such as ICS (He and Saedler, 2005). Hence, structural and functional analysis of *cis*-regulatory elements (promoters and first intons) of *MPF2-like* genes will facilitate in better understanding of the differential expression and eventually the sub-functionalization and/or neo-functionalization of these genes.

3.3.1 *MPF2* and *MPF2-like* promoter analysis

Promoters sequences (16 in total) of the *MPF2-like-A* and *MPF2-like-B* genes from all Withania species and of the *MPF2-like-B* gene from *Tubocapsicum anomalum* (see Materials and Methods) were analyzed for DNA-sequence homology and for transcription factor binding motifs using the Mulan package (Ovcharenko et al., 2005). Limited homology is observed between *MPF2-like-A* and *MPF2-like-B* gene promoters of Withania species throughout the sequenced portion (data not shown). Figure 3.19A reveals DNA sequence homology in a 1.2 kb region upstream of the ATG codon from *MPF2-like* genes of *Withania somnifera*, *Tubocapsicum anomalum* and *MPF2* from *Physalis floridana* (He and Saedler, 2005) in pair-wise comparisons. Numerous variants of MADS-domain transcription factor binding motifs (CArG-boxes) are found in the various promoters but are not shown. Among these, interestingly, a conserved CArG-box is identified in all the promoters except *MPF2-like-B*, which does not express this gene in floral tissues. The position of putative transcription initiation was deduced from the longest cDNA isolated from each species and is given in Fig. 3.19B together with the position of the conserved CArG-box and their separating distance. Figure 3.18C shows DNA sequences of the conserved CArG-box in the various promoters and lists some known MADS-domain proteins recognizing these motifs.

The results revealed by the analysis of each sequence comparison are described here separately. The *MPF2/WSA206* promoters share about 500 bp of homology, most of which, about 400 bp, includes the 5'-UTR region (Fig. 3.19A top panel). Besides the conserved CArG-box, which is a binding site for MADS-box proteins AG, AGL1, SEP1, SEP4 and AGL15 in Arabidopsis (de Folter and Angenent, 2006). Other binding sites for the Homeobox

A)



B)

gene	transcription initiation site	position CArG-box	distance	position of ARF binding site
MPF2	- 472	- 580	108	966
WSA206	- 320	- 430	110	871
TAB201	- 436	- 560	124	1394
WSB206	- 415	no	no	608

C)

WSB206	ggTccc	C	a	A	a	A	A	**g**	t	t	**a**	T	tAtAGTG
WSA206	TTTTTA	C	C	A	T	A	A	A	A	A	G	T	AAAGGTG
MPF2	TTTTTA	C	C	A	T	A	A	A	A	A	G	T	AAAGGTG
TAB201	TTTTTA	C	C	A	T	A	A	A	A	A	G	T	AAAAGTG

AG, AGL1, SEP1
SEP4, AGL15

Figure 3.19: *MPF2-like* promoter and first intron analyses. A) Pair-wise DNA-sequence comparisons using the Mulan Package Online are depicted. Sequence homologies are shown in a window of 50-100%. The first base of ATG is set as 1. The first -1200 bp of each promoter are shown. The bars of different colours represent DNA-binding motifs of the transcription factors indicated. The first 1500 bp from ATG are given for each 1st intron. ARF (Auxin response factor) is recognised by auxin response factor elements (AuxREs). B) shows the position of transcription initiation, the conserved CArG-boxes and the distance between them. Also the distance of the ARF binding site from the ATG is mentioned. The sequences of the conserved MADS-domain transcription factor binding motifs (CArG-boxes) are listed. C) AG, AGL1, SEP1, SEP4 and AGL15 (de Folter and Angenent, 2006) recognize the sequence boxed in solid line.

proteins, AtHB1 and AtHB5, and for the AGP1 and P-binding protein are conserved in all the four promoters tested (Fig.3.19A). Remarkably, all these latter transcription factor-binding sites are located within the 5'-UTR region, while the conserved CArG-box is up-stream of the transcription initiation site.

The *TAB201/WSA206* comparison reveals that these two promoters also share a similar amount of sequence homology (Fig. 3.19A, 2. from top) and the same conserved DNA-sequence motifs are observed here as in the previous case.

The comparison of the *MPF2-like-A* and *MPF2-like-B* gene promoter of Withania *WSA206/ WSB206* (3. from top) shows homology limited to about 300 bp and reveals only the conserved DNA-binding motifs boxed in the 5'-UTR region and not the conserved MADS-domain transcription factor binding motif characteristic of all the other promoters.

This is supported by the comparison of the *MPF2-like-B* genes *TAB201/WSB206* of Tubocapsicum and Withania shown in Fig. 3.19A (bottom panel). Even though the homology extends throughout the entire 5'-UTR region, no conserved MADS-domain transcription factor binding site is detected in this comparison.

While transcription initiation sites and the position of the conserved CArG-box vary in the different promoters, the distance separating them seems to be fairly constant (Fig. 3.19B). In the corresponding region of the *WSB206* promoter, however, no defined CArG-box motif is found (Fig. 3.19C). This could suggest that tissue-specific differences in expression of *MPF2-like-A* and *MPF2-like-B* genes in Withania are attributed to presence and/or absence of CArG-motif. While *MPF2*, *WSA206* and *TAB201* are expressed in all tissues tested, the *MPF2-like-B* gene (*WSB206*) of Withania, is not (Fig. 3.15). Whether this is due to the missing conserved CArG-box in its promoter region remains to be tested.

3.3.2 *MPF2* and *MPF2-like* first intron analysis

Like promoters the 1st introns (16 sequences in total) from *MPF2-like-A* and *MPF2-like-B* genes of all the Withania and Tubocapsicum species were analyzed using sequence homology and binding motifs in Mulan package (Ovcharenko et al., 2005). Patches of homology are found in all the sequences tested (data not shown). Here also many CArG-boxes are found scattered along the entire length of the intron. Figure 3.19A depicts sequence homology in 1.5 kb down-stream of ATG. Interestingly, a 100 bp highly conserved region in all the *MPF2-like* 1st intron sequence comparison has been ascertained at different positions. When this region is further narrowed down, the auxin response factor (ARF) elements (AuxREs) are detected in all intron sequences except *WSA206* 1st intron. Figure 3.19B shows the distance of ARF from

Serial Nr.	%age +tive	Construct name	Promoter length, CArG-box position, transcription initiation site		Intron length, CArG-box position, ARF position		Leaf	Pedic-el	Recep-tacle	Sepals	Petals	Stamens	Carpels	Silique
1	27	WS a-Pl+I	cbox-2050, cbox-2550,	2606	ARF691,cbox476,cbox2081	2373	+	-	+	-	-	++	-	+tip
2	41	WS a-Pl		2606		-	+++	+	+	++++	+	+++	++++	++++
3	45	WS a-Pls(cbox)	Transcription-ini -320,	2066		-	++	-	-	-	-	++	-	-
4	10	WS a-Ps+IARF		1261		712	+	-	++	-	-	+	+	+
5	9	WS a-Ps+ImARF		1261		712	-	-	++	-	-	++	-	-
6	88	WS a-Ps		1261		-	++++	+	+	-	-	+++	+++	+++
7	60	WS a-Pss		560		-	++++	++	+++	-	-	+++	+++	+++
8	0	WS a-Pss+IARF		560		712	-	-	-	-	-	-	-	-
9	28	WS a-Pss+ImARF		560		712	++	++	-	+++	-	-	+	-
10	3	WS b-Pl+I	cbox-698, cbox-1754,	3129	ARF 428, cbox1133,1627,	1827	-	-	+++	-	++	+++	+	+
11	98	WS b-Pl		3129		-	+++	++++	+++++	-	++	+++	+++++	+++++
12	96	WS b-Pls(cbox)	Transcription-ini -415,	750		-	-	+++	+++	++++	++	++++	+++	+++
13	0	WS b-Ps+IARF		1269		458	-	-	-	-	-	-	-	-
14	39	WS b-Ps+ImARF		1269		458	+++	+++++	+++++	+++++	+++++	+++++	+++++	+++++
15	33	WS b-Ps		1269		-	+++	-	+	-	-	+	-	+
16	31	WS b-Pss		612		-	+++	-	+++	-	-	+	-	-
17	33	WS_b-Pss+IARF		612		458	+++	+++	+	-	-	++	-	+tip
18	0	WS b-Pss+ImARF		612		458	-	-	-	-	-	-	-	-
19	24	TA b-Pl+I	cbox-1290,cgbox-1557	2294	ARF1214,cbox697,1493,210	3704	-	++++	++++	++++	++	+++	++++	++++
20	90	TA b-Pl	cbox-1851,	2294	cbox 2888,2995,3048	-	+++	++++	++++	-	-	+++	++++	++++
21	32	TA_b-Ps(cbox)	Transcription-ini -436,	1282		-	+++	++	+++	-	+	++	-	-
22	50	TA b-Ps+IARF		1398		1237	-	-	-	-	-	+	-	-
23	75	TA b-Ps+ImARF		1398		1237	++	++	++	+	-	+	+	+
24	26	TA b-Ps		1398		-	++	++	++	-	-	++	-	-
25	80	TA_b-Pss		554		-	+++++	-	++	+++++	++	++++	+++++	+++
26	0	TA b-Pss+IARF		554		1237	-		-	-	-	-	-	-
27	20	TA b-Pss+ImARF		554		1237	++	++	++	-	-	++	-	+tip
28	79	MPF2-Pl+I	cbox-1784,-2091,cg-1488,	2232	ARF786,cbox549,855,1429,	2165	-	++	++++	-	+	++++	+	+tip
29	32	MPF2-Ps+IARF	Transcription-ini -472,	1243		811	-	+	++	-	-	+++	++	+
30	50	MPF2-Ps+ImARF		1243		811	+++	-	+++	-	-	+++	+++	+++
31	72	MPF2-Ps		1243		-	++	-	++	+++	-	+++	-	+
32	21	MPF2-Pss		576		-	-	-	-	-	-	++	-	-
33	28	MPF2-Pss+IARF		576		811	-	-	++	-	-	++	-	-
34	23	MPF2-Pss+ImARF		576		811	++	-	-	-	-	-	-	-
35	9	V001-1-Pl+I	cbox-440,-836,-1285,	1950	ARF1344,cbox186,812,945,	3444	-	-	-	-	-	+	-	-
36	0	V001-1-Pls(cbox)		1950	cbox2377,2539,2781,2896,3311	-	-	-	-	-	-	-	-	-
37	17	V001-1-Pls(cbox)	Transcription-ini -394	809		-	-	-	-	+++	++	+++	-	-
38	81	V001-2-Pl		2071		-	++	-	-	+++++	++++	+++++	++++	++++
39	31	V001-2-Pls(cbox)		810		-	+	-	-	++	++	+++	+	-

Table 3.5: Specifications of promoter and intron:GUS constructs and overview of expression patterns in Arabidopsis. WS_a= *WSA206*; WS_b= *WSB206*; TA_b= *TAB201*; V001-1= Vassobia 1st sequence; Vassobia 2nd sequence; P=promoter; l= long; I= 1st exon-1st intron; s= short; ss= very short; m= mutated;+++++= very strong; ++++= strong; +++= moderate; ++= weak; += very weak; -= no expression; tip= silique tip.

ATG which is variable in all the sequences. It can be speculated that ARF plays a regulatory role in the leaf development.

In addition, INDETERMINATE for flowering (ID1), transcription factor binding sites are found at different positions both in the introns as well as in the promoters. ID1 has one binding site in *WSA206*, five in *WSB206*, two in *TAB201* and one in *MPF2* (data not shown). As ID1 is usually involved in flowering time regulation, hence, role of *MPF2-like* genes in flowering time control can be assumed.

Based on Mulan analyses, the functionality of promoters and first introns in regulating the *MPF2-like* genes expression in the leaves as well as in the flowers was evaluated in transgenic Arabidopsis stably and in *Nicotiana benthamiana* transiently.

3.3.3 Strong GUS signals are detected in flowers of *MPF2-like-B* promoter:GUS transgenic Arabidopsis plants

A complete overview of the *MPF2-like* promoter and intron GUS expression analyses in Arabidopsis is given in table 3.5. Generally, most of the tansgenic Arabidopsis plants harbouring constructs demonstrated GUS expression patterns in leaves and inflorescences

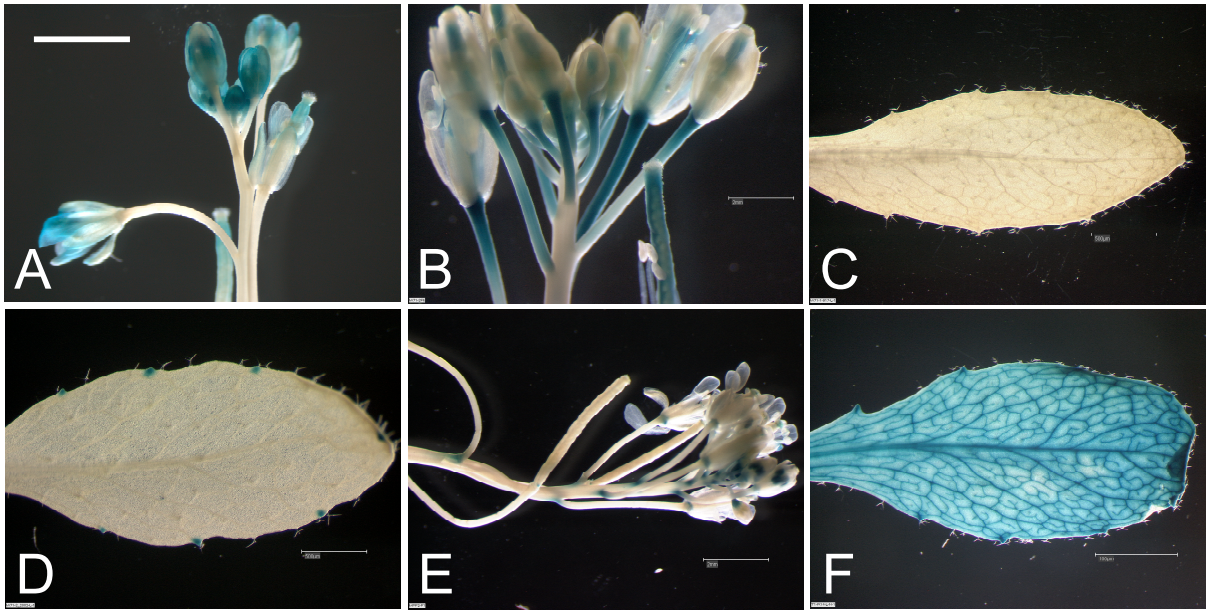


Figure 3.20: Histochemical GUS staining assay of transgenic Arabidopsis leaves and inflorescences

A) WS_a-Pl+I (*WSA206*-promoter long + 1st exon-1st intron); B) WS_b-Pl (*WSB206*-promoter long); C) WS_b-Pl+I (*WSB206*-promoter long + 1st exon-1st intron); D) WS_a-Pl+I (*WSA206*-promoter long + 1st exon-1st intron); E) MPF2-Pl+I (*MPF2*-promoter long + 1st exon-1st intron); F) TA_b-Pss (*TAB201*-promoter very short). For details see Table 3.5. Scale bar = 2mm.

including receptacle, stamens and carpels. Stamens were the organs where most of the promoter and intron constructs exhibited strong GUS expression.

Although there were obvious differences in the expression patterns of GUS gene in the leaves tissues (Fig. 3.20 C,D). However, the most striking observation was the detection of very strong GUS expression in the inflorescence of *MPF2-like-B* genes (compare Fig. 3.20 A,B and construct number 2 and 11 in Table 3.5). Floral expression of *MPF2-like-A* genes is compatible with expression found in native plant, but expression of *MPF2-like-B* genes in the flower is not anticipated.

3.3.4 First introns generally suppress the *MPF2-like* expression in the leaf

Generally very obvious GUS expression patterns were detected in the inflorescence if long promoters with 1stexon-1st introns were transformed into Arabidopsis (Fig. 3.20 E and table 3.5, constructs number 1, 10, 19, 28). The expression was mainly concentrated in the receptacle and stamens. Remarkably, no leaf expression was visible except WS_a- Pl+I (*WSA206*-promoter long + 1st exon-1st intron) construct as shown in the Fig. 3.20D and construct number 1 in the table 3.5. It can be inferred that *MPF2-like* gene expression is suppressed in the leaf by 1st intron. On the other hand, the plants transformed with *MPF2-like* promoter:GUS constructs only, exhibited GUS patterns in almost all the tissues including leaves and flowers (Fig. 3.20 A,B,F). Receptacle, stamens and carpels were the tissues where maximum expression was visible.

3.3.5 *MPF2-like* first introns suppress the YFP signals in transiently transformed *Nicotiana benthamiana* leaves

In most of the *MPF2-like* peomoter as well as intron constructs, YFP signals were detectable in the nucleus (data not shown). It was ascertained that the YFP constructs carrying intron (promoter long + 1stexon-1st intron), showed low YFP signal intensity (Fig. 3.21D). However, with promoter:YFP constructs strong YFP signals could be detected in the *N. benthamiana* leaf cells (Fig. 3.21C). This indicate that 1st introns are suppressing the YFP reporter gene driven under 35 promoter in the leaves. These results are also supported by GUS expression patterns in the transgenic Arabidopsis.

What can be concluded from above observations is that 1st introns suppress the expression of *MPF2-like* genes in leaves.

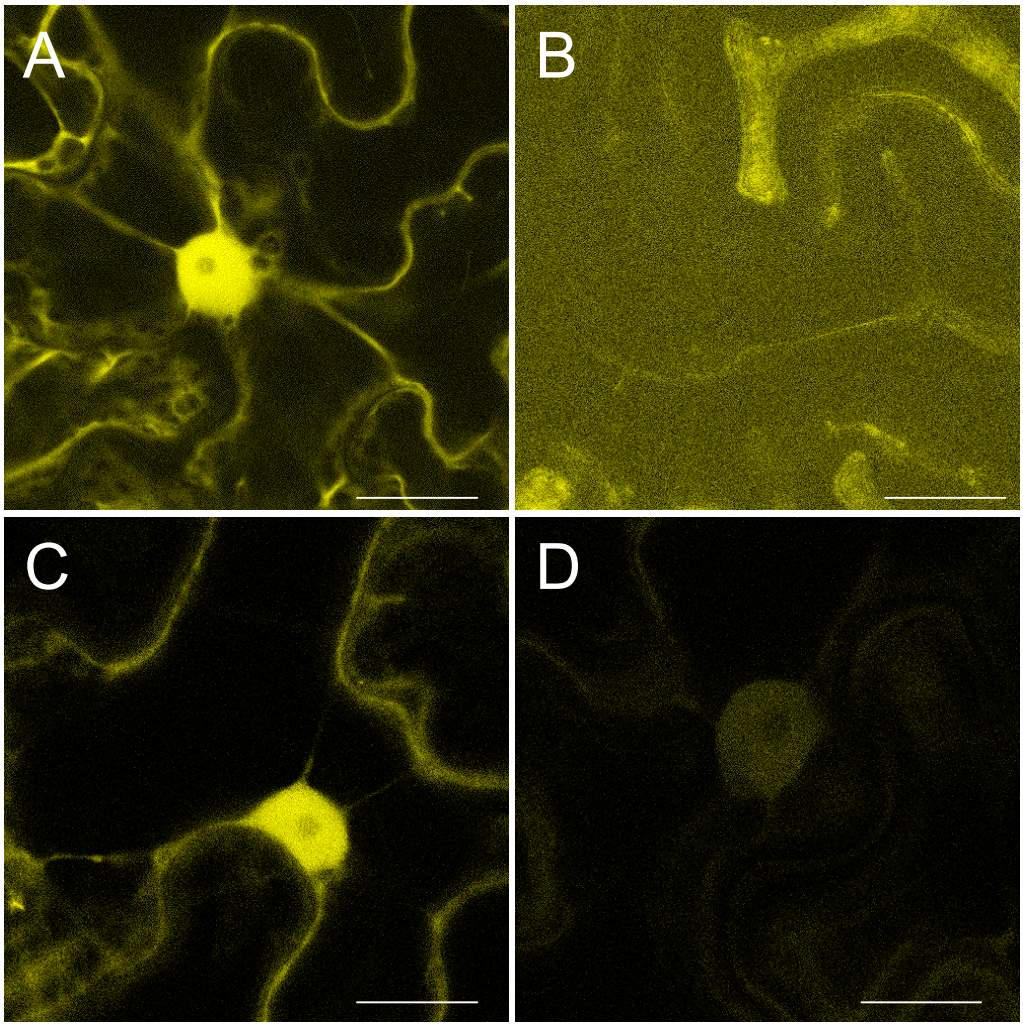


Figure 3.21: CLSM images showing YFP expression in cells of transiently transformed *N. benthamiana* leaves

A) Positive control (*35S:YFP*); B) Negative control (YFP only); C) Promoter:YFP for *WSA206* gene; D) Promoter long + 1st exon-1st intron:YFP for *WSA206* gene. Scale bar = 10 μm.

Differentiation in regulation and functions of *MPF2-like* genes can be studied in more details by finding their interacting partners in yeast two hybrid analysis.

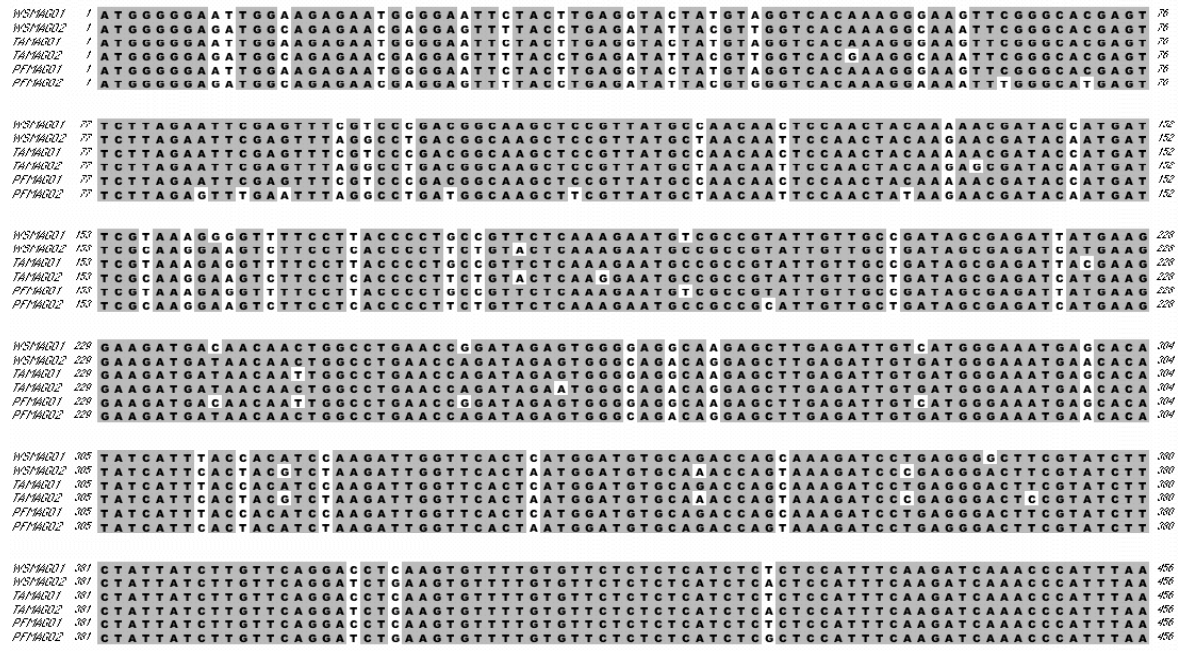
3.4 Yeast two-hybrid analysis

The structural and functional characterization of MPF2-like proteins demonstrated that after duplication and diversification, these genes are well on their way probably, to sub-functionalization and/or neo-functionalization fate. Differentiation in functions can be elaborated in more details by finding their interacting partners. Therefore, MPF2-like and MAGO NASHI-like proteins from Withania, Tubocapsicum and Physalis were tested for their interaction with each other in yeast two-hybrid system. Furthermore, Arabidopsis oligo-dT cDNA yeast library screening and confirmation with MPF2-like proteins was undertaken.

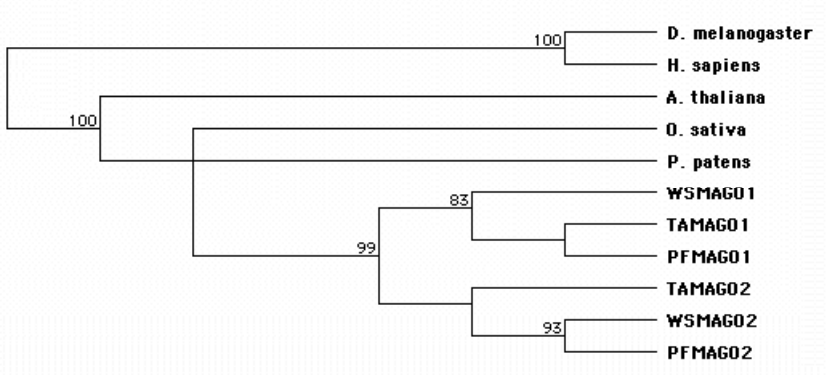
3.4.1 Molecular characterization of Withania and Tubocapsicum MAGO NASHI-like proteins

Coding sequences of *MAGO NASHI-like* genes were amplified from cDNA of Withania and Tubocapsicum with primers based on *PFMAGO1* and *PFMAGO2* gene sequences from *Physalis floridana* (He et al., 2007). Although many clones of *MAGO NASHI-like* genes were sequenced, only two types of gene sequences could be identified for *W. somnifera* and *T. anomalum*. These are named *WSMAGO1* and *WSMAGO2* for *W. somnifera* and *TAMAGO1* and *TAMAGO2* for *T. anomalum*, respectively. Figure 3.22 shows the ClustalW alignments of *MAGO NASHI-like* genes. Both *WSMAGO1* and *WSMAGO2* share a coding sequence of 453 nucleotides. Coding sequences are 88% similar (50 nucleotide differences), while protein sequences are 96% identical (5 amino acids are different in a total of 151). In Tubocapsicum, *TAMAGOs* share a homology of 88% (51 nucleotide differences in 453) at nucleotide level and 94% (8 aa are different in 151) at amino acid level. This sequence comparison suggests that both the Withania and Tubocapsicum *MAGO NASHI-like* genes have the same tendency of homology conservation. *WSMAGO1* and *TAMAGO1* are 98% similar at both the nucleotide as well as protein level. An identical degree of conservation is also observed for the second copy *WSMAGO2* and *TAMAGO2*. Surprisingly, a homology of 96% at the nucleotide level and 100% at the amino acid level is noticed when *WSMAGO2* and *PFMAGO2* are compared. This high conservation indicates that both proteins fulfil similar functions. This is also evident from the clustering of WSMAGO1, TAMAGO1, and PFMAGO1, separately from WSMAGO2, TAMAGO2 and PFMAGO2. In Figure 3.22B a simple UPGMA tree (Unweighted Pair Group Method with Arithmetic Mean) shows that all the proteins of Solanaceous species are clustered together. Apart from proteins of Tubocapsicum,

A)



B)



C)

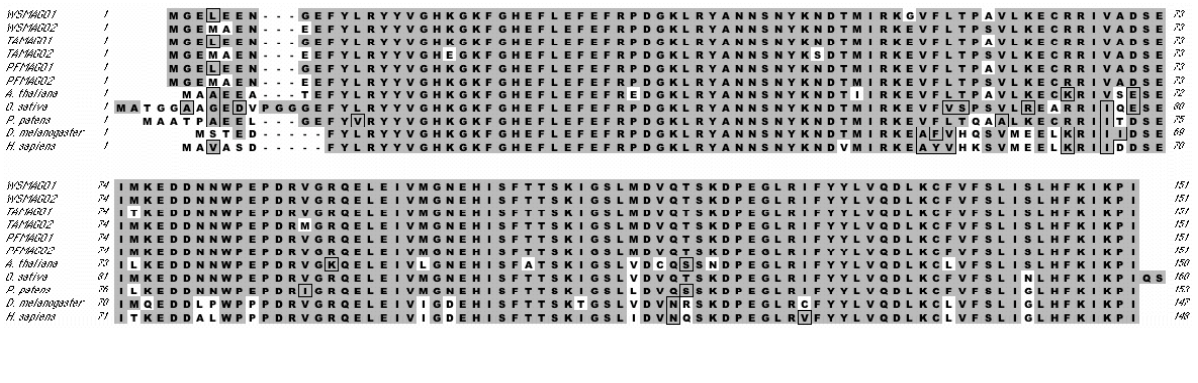


Figure 3.22: Sequence analyses of *MAGO NASHI-like* genes. A) ClustalW nucleotide alignments of Withania, Tubocapsicum and Physalis *MAGO NASH1-like* homologs. B) UPGMA (Unweighted Pair Group Method with Arithmetic Mean) phylogenetic tree based on amino acid sequences from different species in Mac Vector™ 7.2.3 programme. Numbers above the branches indicate the bootstrap values of 1000 replicates. C) ClustalW amino acid alignments of MAGO MASHI-like proteins from different species.

WSMAGOs share 70-98% identity with other homologs in different species like Physalis, rice, moss and animals (Fig. 3.22C). Variability is more prominent in the N-terminal regions than others. This high conservation in such a diverse range of species foreshadows that MAGO NASHI-like proteins have conserved function.

3.4.2 R-motif at the C-terminus of WSA206 protein represses its self-activation

Full-length cDNAs encoding MPF2-like and MAGO NASHI-like proteins were introduced into yeast 2-hybrid prey and bait vectors (see Materials and Methods), to check their intrinsic transcriptional activating properties and its potential toxicity in yeast. All transformed yeasts grew normally on tryptophan-deficient (bait) or leucine-deficient (prey) plates, indicating that transgene is not toxic to yeast. No growth on histidine-deficient plates (SD/-Leu-Trp-His) indicated that WSA206, WSB206 and TAB201 could not auto-activate. But MPF2 (a close homolog of MPF2-like) can auto-activate due to its proline rich acidic C-terminal domain. When this domain is removed from MPF2, the ability to self-activate no longer exists (He et al., 2007). The C-terminus of all the MPF2-like proteins is also proline rich and acidic except third last proline residue (Fig. 3.23A,B). Surprisingly, when the last 8aa residues from WSA206 were removed to generate WSA206ΔC246, the truncated version could self-activate. This suggests that the 8aa extension is involved in the repression of self-activation. Here this 8 aa extension is named as R-motif (Fig. 3. 23A,B). Interestingly, when this R-motif was added to the C-terminus of WSB206 to make WSB206∇C257, this extended version, contrary to WSA206ΔC246, did exhibited properties of self atcivating HIS3 reporter gene on its own like in original WSB206 case. From this experiment it can be inferred that the C-terminus of these proteins is essential for their functioning. Any alteration in this region is reflected in the changed behaviour of protein. The change of function in terms of self-activation of MPF2-like proteins may be attributed to conformational changes in the proteins after a deletion or addition at the C-terminus.

Parallel to auto-activation, auto-binding was also tested for MPF2-like proteins. No case of self-binding in any protein was observed except for WSB206∇C257. This protein demonstrated blue colour in β-galactosidase assay (Fig. 3. 23B).

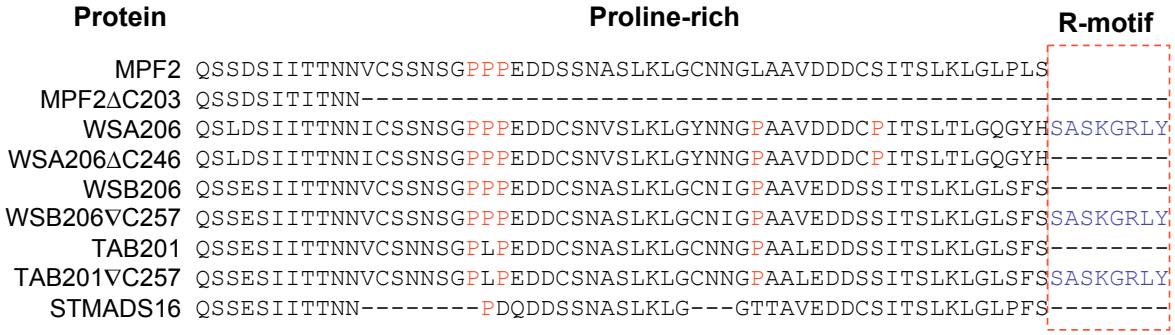
3.4.3 MPF2-like-A proteins can form homodimers

The ability of MPF2-like proteins to form homo or heterodimers in yeast was investigated. All the MPF2-like-A (WSA206*)* proteins could form homo-dimers (Fig. 3.23B and 3.24A,B). The ability to form homodimers was still retained by both the WSA206ΔC246 and

WSB206∇C257 proteins even after truncation and addition, although, their self-activating properties changed (Table 3.6).

In addition to homodimer formation these proteins could form heterodimers with other proteins. TAB201 showed the ability to interact with both the Withania proteins. In table 3.6 all the experiments for testing auto-activation, dimer formation and interactions related to MPF2-like and MAGO NASHI-like proteins have been summarized. The changed WSA206ΔC246 and WSB206∇C257 proteins exhibit interactions with WSA206, WSB206 and TAB201. Moreover, there is a trend for directional interaction in these proteins. From the figure 3.24A and B, it is apparent that horizontal interactions are more dominant than vertical ones.

A)



B)

Protein	Self-activation A	Self-activation B	Self-binding	Homodimer formation
MPF2	-	-	nt	+
MPF2ΔC203	+	+	nt	-
WSA206	-	-	-	+
WSA206ΔC246	+	+	-	+
WSB206	-	-	-	-
WSB206∇C257	-	-	+	+
TAB201	-	-	-	-
TAB201∇C257	-	-	-	+
STMADS16	-	-	nt	+

Figure 3.23: Interacting characteristics of MPF2 and MPF2-like proteins are determined by differences in their C-termini. A) The gaps were introduced to optimize the alignment of conserved residues. The prolines in the proline-rich domain are shown in red letters. The R-motif is shown in blue letters and a blue dotted line indicates the removal of 8 aa from WSA206. B) The ability to self-activate and self-bind in the yeast 2-hybrid

system was revealed by cell growth on SD/-Trp-Leu-His is indicated by letter A whereas letter B indicates expression of ß-galactosidase, as determined by the development of a blue colour in a non-lethal colony assay. Symbols + or – indicate growth or no growth or the blue or white phenotype, respectively; nt means not tested. Homodimer formation (+) was detected as described in Results. Note the self-activation of WSA206ΔC246 in and self-binding WSB206∇C257 in B, respectively.

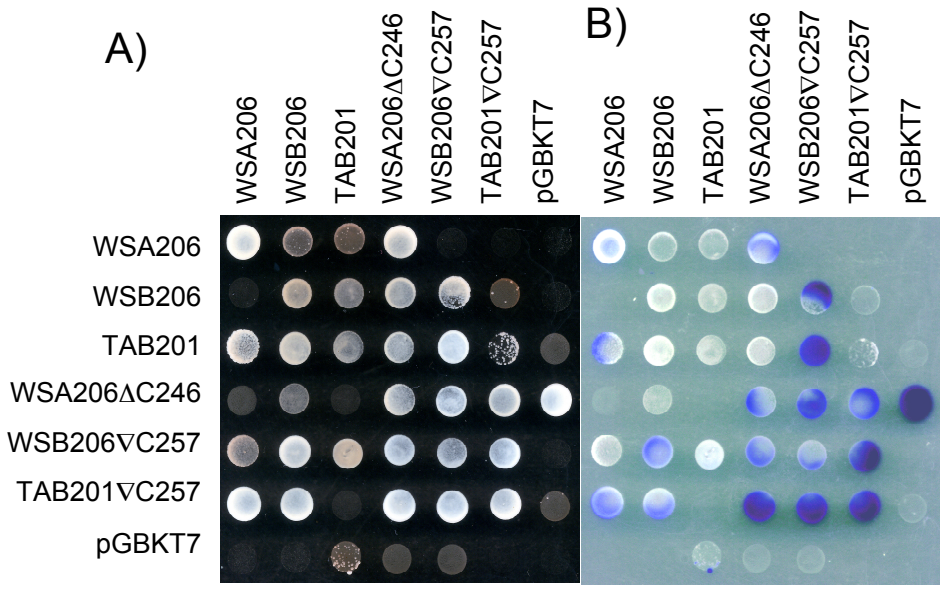


Figure 3.24: Confirmation of MPF2-like interactions. The yeast interaction matrix is shown in both the figures here. A) Cell growth in the yeast 2-hybrid system on a high-stringency selective plate (SD/-Trp-His-Leu-Ade and 3.0 mM 3-AT). B) Non-lethal ß-galactosidase assay to confirm the interactions. The expressed prey proteins are indicated above and the bait proteins to the left of the panels.

3.4.4 MAGO NASHI-like proteins interact with MPF2-like-B and not with MPF2-like-A proteins

A total of 6 MAGO NASHI-like protein constructs were used to study their protein interactions in the yeast system. Withania and Tubocapsicum constructs (WSMAGO1, WSMAGO2, TAMAGO1 and TAMAGO2) were made as described in Materials and Methods, two constructs from Physalis (PFMAGO1 and PFMAGO2) were kindly provided by Chaoying He. Analysis of yeast transformants revealed that neither of the MAGO NASHI homologs could activate HIS3 reporter genes on its own and neither was toxic to yeast (Table 3.6). Thus, all the proteins could not show auto-activation or auto-binding. This analogous behaviour of MAGO NASHI-like proteins might be attributed to their highly similar C-terminus, which is thought to be involved in self-activation. In MAGO NASHI-like proteins

as already mentioned, significant variability is found at the N-terminus. These results are in agreement with He et al., (2007).

Table 3.6 shows that all the MAGO NASHI-like proteins are unable to homo-dimerize. However, with MPF2-like proteins these proteins do interact. Most of the MAGO NASHI-like proteins show interactions with WSB206 proteins except TAMAGO1. TAB201 retains ability to interact with WSMAGO2 and TAMAGO2 but it shows no affinities for PFMAGO2 although, WSMAGO2 and PFMAGO2 proteins are highly conserved. Physalis MAGO NASHI-like proteins interact more strongly with WSB206 than Withania and Tubocapsicum MAGO NASHI-like proteins, which demonstrate only week interactions.

	WSB206	WSA206	TAB201	WSB206∇C257	WSA206ΔC246	PFMAGO1	PFMAGO2	WSMAGO1	WSMAGO2	TAMAGO1	TAMAGO2	pGADT7
WSB206	+	-	-	+	-	-	-	-	-	-	-	-
WSA206	+	++	-	++	+	-	-	-	-	-	-	-
TAB201	++	++	++	++	++	-	-	-	+	-	+	-
WSB206∇C257	++	++	++	++	++	+	+	++	++	+	++	
WSA206ΔC246	++	++	+	++	++	-	-	+	-	-	-	++
PFMAGO1	++	-	-	+	+	-	-	nt	-	nt	nt	-
PFMAGO2	++	-	-	+	-	-	-	nt	-	nt	nt	-
WSMAGO1	+	-	-	++	-	nt	nt	-	-	-		-
WSMAGO2	+	-	+	++	+	-	-	-	-	-	-	-
TAMAGO1	+	-	-	++	-	nt	nt	-	-	-	-	-
TAMAGO2	-	-	-	++	-	nt	nt	-	-	-	-	-
pGBKT7	-	-	-	+	-	-	-	-	-	-	-	-

Table 3.6: Reciprocal interactions among MPF2-like and MAGO NASHI-like proteins. ++: Very strong interactions; +: interactions; -: no interactions; nt: not tested. The expressed prey proteins are indicated above

and the bait proteins to the left of the table. Red box highlights the auto-activation of WSA206ΔC246 and protein after truncation while blue box indicates auto-binding of WSB206∇C257 after extension.

However, no interactions between WSA206 and all the MAGO NASHI-like proteins are observed. Similarly, WSB206∇C257-proteins exhibit stronger interactions for all the MAGO NASHI-like proteins than WSA206ΔC246, which interacts with Withania MAGO NASHI-like proteins and Physalis MAGO1. But these interactions are ambiguous due to the fact that WSB206∇C257 and WSA206ΔC246 can seld-bind self-activate, respectively (Table 3.6). Among each other, MAGO NASHI-like proteins tested so far, seldom interact. Here again the trend for horizontal interactions is more dominant than vertical ones.
Arabidopsis cDNA oligo-dT yeast library screening with MPF2-like-proteins is employed to differentiate the functions of MPF2-like proteins.

3.4.5 Screening of Arabidopsis oligo-dT cDNA yeast library

The functional divergence between MPF2-like proteins was further elucidated by finding out their interactors in Arabidopsis. An Arabidopsis oligo-dT cDNA yeast library was screened with full length WSA206, WSB206 and TAB201 as bait in the yeast 2-hybrid system (for details see Materials and Methods). For each protein more than 200 interactors were sequenced and the BLAST search showed multiple hits for a variety of interactors. Here only MADS-domain interactors are taken into account. The results obtained are described separately for each bait.
For WSA206 proteins a total of 70 positive colonies obtained from Arabidopsis library were analyzed further, leading to the identification of 9 different MADS-domain proteins: SOC1, AGL27, AGL19, AGL31, AGL42, AGL68, MAF4, SEP3 and SEP4 (Table 3.7). All these proteins are known to play a role in either flower development or flowering time control. They include some, but not all, of the proteins known to interact with AGL24 (de Folter et al., 2005). AGL24 is an ortholog of MPF2-like in Arabidopsis. The failure to isolate, for example, AP1 and SEP1, in this screen might have been a consequence of the relatively small number of colonies examined in detail or of the heterologous nature of the system.
Contrary to WSA206 protein, remarkably, no MADS-domain interactors could be fished out when WSB206 protein was used as bait although more than 300 clones were sequenced.
However, TAB201 showed an intermediate behaviour between WSA206 and WSB206 by interacting with three MADS-domain proteins, which were SOC1, MAF5 and AGL27.

Bait protein	Interactor	Hits
WSA206	SEP4 (SEPALLATA4); DNA binding / transcription factor (AGL3)	6
	SEP3 (SEPALLATA4); DNA binding / transcription factor (AGL9)	2
	AGL27, MAF1 (MADS AFFECTING FLOWERING 1)	21
	AGL19 (AGAMOUS-LIKE 19); transcription factor	7
	AGL31 (AGAMOUS LIKE MADS-BOX PROTEIN 31) MAF3	12
	AGL42 (AGAMOUS LIKE 42); transcription factor	8
	AGL20 (AGAMOUS-LIKE 20); transcription factor (SOC1)	7
	MAF4	1
	MAF5, AGL68	3
	Auxin-responsive family protein	3
	AKIN10 (ARABIDOPSIS SNF1 KINASE HOMOLOG 10)	14
	AT)SRC2/SRC2 (SOYBEAN GENE)	7
	C2 domain-containing protein (*Arabidopsis thaliana*)	21
	Zinc finger (CCCH-type) family protein (*Arabidopsis thaliana*)	13
	ATSYTA/NTMC2T1.1/NTMC2TYPE1.1/SYTA (*Arabidopsis thaliana*)	7
	Multicatalytic endopeptidase complex alpha chain	21
	Calnexin 1 (CNX1) (*Arabidopsis thaliana*)	4
	EMB1507 (EMBRYO DEFECTIVE 1507)	5
	PAD1 (20S PROTEASOME ALPHA SUBUNIT PAD1)	18
	Unknown Arabidopsis proteins	46
WSB206	IAA16 (indoleacetic acid-induced protein 16); transcription factor	4
	AXR4 (AUXIN RESISTANT 4)	3
	RAP2.4 (related to AP2 4); DNA binding / transcription factor	5
	EMB1417 (EMBRYO DEFECTIVE 1417)	19
	ATSIP2 (ARABIDOPSIS THALIANA SEED IMBIBITION 2); hydrolase	17
	RCD1 (RADICAL-INDUCED CELL DEATH1)	2
	AKIN10 (ARABIDOPSIS SNF1 KINASE HOMOLOG 10)	13
	MYB4 (myb domain protein 4); transcription factor	5
	Zinc finger (CCCH-type) family protein	14
	DnaJ protein	18
	Immunophilin, putative / FKBP-type	4
	MYB3 (myb domain protein 3); DNA binding	3
	40S ribosomal protein S20 (RPS20C) (*Arabidopsis thaliana*)	11
	C2 domain-containing protein (*Arabidopsis thaliana*)	10
	Putative RNA-binding protein (*Arabidopsis thaliana*)	2
	FPF1 (FLOWERING PROMOTING FACTOR 1, *Arabidopsis thaliana*)	2
	Ferredoxin-related (*Arabidopsis thaliana*)	3
	PAD1 (20S PROTEASOME ALPHA SUBUNIT PAD1)	23
	AJH1/CSN5A/JAB1 (COP9 SIGNALOSOME 5A) (*Arabidopsis thaliana*)	13
	Leucine-rich repeat family protein (*Arabidopsis thaliana*)	3
	Unknown Arabidopsis proteins	59
TAB201	AGL20 (AGAMOUS-LIKE 20); transcription factor (SOC1)	1
	AGL27, MAF1 (MADS AFFECTING FLOWERING 1)	4
	MAF5, AGL68	3
	C2 domain-containing protein (*Arabidopsis thaliana*)	23
	AKIN10 (ARABIDOPSIS SNF1 KINASE HOMOLOG 10)	11
	EMB2726 (EMBRYO DEFECTIVE 2726)	18
	ATSYTA/NTMC2T1.1/NTMC2TYPE1.1/SYTA (*Arabidopsis thaliana*)	13
	pEARLI4 (*Arabidopsis thaliana*])	2
	late embryogenesis abundant family protein / LEA family protein	3
	MYB4 (myb domain protein 4); transcription factor	6
	PAD1 (20S PROTEASOME ALPHA SUBUNIT PAD1)	23
	DPA; transcription factor (Arabidopsis thaliana)	3
	NTMC2T4/NTMC2TYPE4; lipid binding (*Arabidopsis thaliana*)	13
	MYB3 (myb domain protein 3); DNA binding	5
	40S ribosomal protein S20 (RPS20C) (*Arabidopsis thaliana*)	24
	protein phosphatase 2C, putative / PP2C	7
	nuclear transport factor 2 (NTF2) family protein	3
	putative ATP-dependent RNA helicase (*Arabidopsis thaliana*)	4
	DPA; transcription factor (*Arabidopsis thaliana*)	2
	Unknown Arabidopsis proteins	10
		43

Table 3.7: List of Arabidopsis interactors for MPF2-like proteins (WSA206, WSB206, TAB201) used as baits. Interactors highlighted as red are the MADS- domain interactors while blue indicates the interactors, which are common to all the three, bait proteins used. Note the non-existence of MADS-domain interactors for WSB206 bait protein.

3.4.6 Non-MADS-domain interactors

As mentioned above, MADS-domain proteins can also interact with non-MADS-domain proteins, and indeed, in addition to the MADS-domain clones described in the previous section, a large number of non-MADS-domain interactors were found with all the three proteins used as baits. Among them some important included: auxin-responsive family protein (ARF); indoleacetic acid-induced protein (IAA16); EMBRYO DEFECTIVE (EMB1417); ARABIDOPSIS THALIANA SEED IMBIBITION (ATSIP2); C2-domain containing protein; MYB3; MYB4; Zinc finger (CCCH-type) family protein; AKIN; PADI. Some members of RNA-binding proteins were also found but majority of the interactors were Arabidopsis proteins with unknown functions. The interaction of MPF2-like proteins with these overlapping non-MADS-domain interactors such as AKIN10, C2 domain containing protein, PADI, EMBRYO DEFECTIVE and unknown Arabidopsis proteins predict their common functions which may be normal development of plant body parts including leaf flower and seed development.

3.4.7 Verification of interactions in yeast

In order to confirm the interactions identified in the yeast library screens and some other proteins involved in flower development, both the bait and prey constructs for MADS-domain interactors (AP1, SEP1, SEP2, SEP3, SEP4 and SOC1) were independently transformed into the yeast strain AH109. After cotransformation, cells were plated on SD/- Trp-Leu-His-Ade, plus 3.0 mM 3-AT. Transformed cells carrying these bidirectional combinations were recovered (Fig. 3.25A), and when subjected to the non-lethal ß-galactosidase assay these cells turned blue (Fig. 3.25B), thus confirming the results of library screen. The confirmations of interactions of MPF2-like proteins with each other have been already described in the previous section. Fig. 3.25 shows that WSB206 shows weak interactions with SEP4 but WSA206 is able to interact strongly with SEP3, SEP4 and SOC1, which were fished out from Arabidopsis yeast library screen as well. In addition, WSA206 shows affinities for AP1 and SEP1 but not SEP2. TAB201 exhibits interaction with SEP1, SEP3 and SEP4 like WSA206 although these proteins could not be fished out from library screen. TAB201 interacts very weakly with SOC1, which was found as an interacting protein in the library screen. Arabidopsis proteins can also interact with each other with specificities. The interaction patterns obtained for Arabidopsis proteins among themselves are in agreement with results of de Folter et al., (2005) except in the cases of SEP1ΔC168 interaction with SEP4 and SEP2ΔC168 with SEP1. However, these anomalies are beyond the scope of this study.

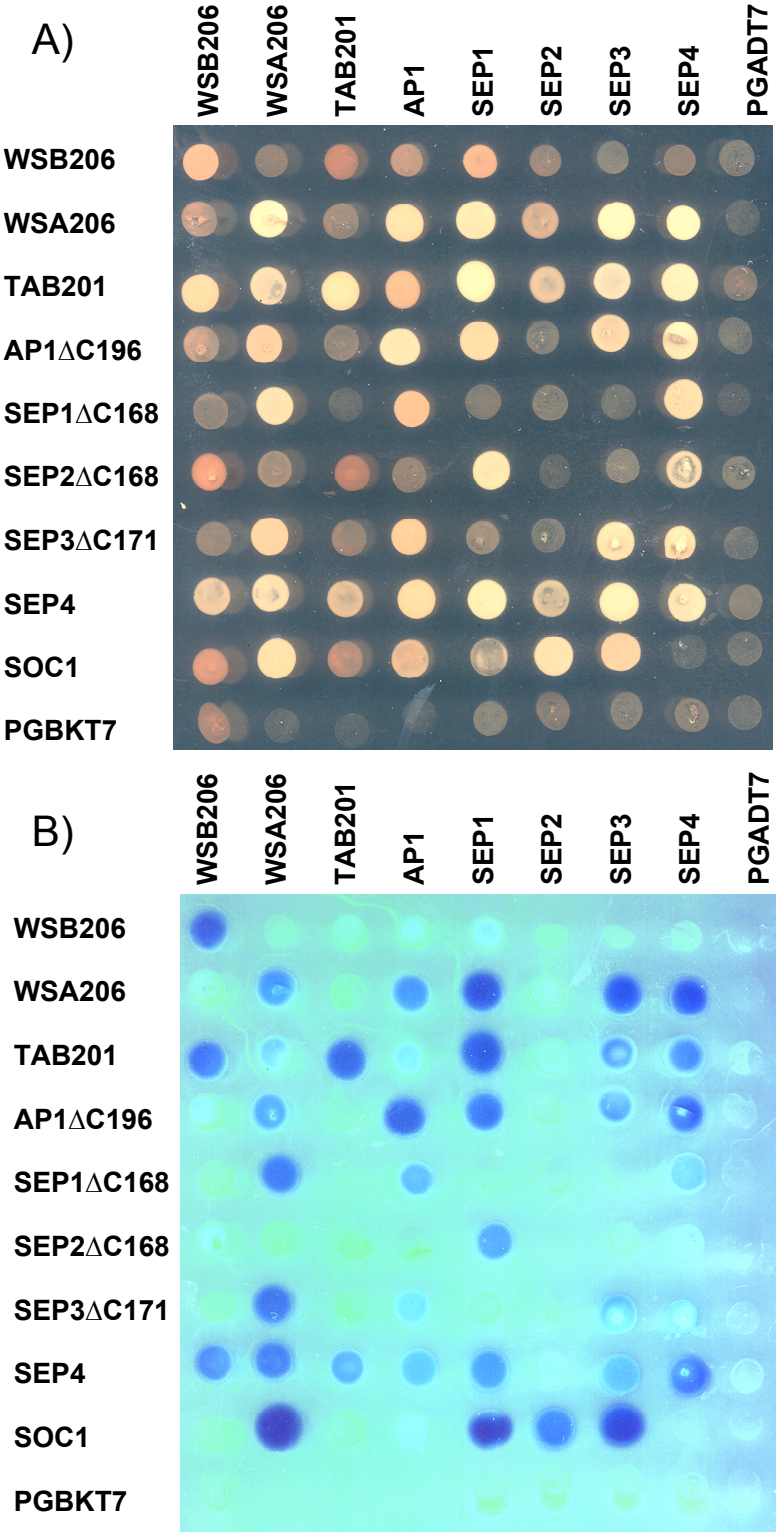


Figure 3.25: Confirmation of MPF2-like and selected Arabidopsis proteins interactions. A) The interaction matrix is shown here for Solanaceous and Arabidopsis MADS-domain proteins. Cell growth in the yeast 2-hybrid system on a high-stringency selective plate (SD/-Trp-His-Leu-Ade and 3.0 mM 3-AT). B) Non-lethal ß-galactosidase test to confirm the interactions. The expressed prey proteins are indicated above and the bait proteins to the left of the panels. Note the interaction of WSA206 with MADS- domain proteins.

4 Discussion

The previously reported results explicitly demonstrate the diversity of inflated calyx syndrome (ICS) and role of plant hormones in the formation of ICS in the Withaninae. The protein structures of the two MADS-domain proteins, MPF2-like-A and MPF2-like-B - duplicated probably, due to allopolyploidization - have been elucidated through phylogentic and functional analyses. Furthermore, the results obtained from expression studies in transgenic *Arabidopsis thaliana* subsume the regulatory functions of *MPF2-like* promoter and first intron whereas the interactions of MPF2-like proteins with other MADS-domain as well as non-MADS-domain proteins are revealed by yeast two-hybrid analysis. Based on computational and experimental data *MPF2-like* gene duplication, gene loss, diversification, sub-functionalization and/or neo-functionalization and their contribution to ICS formation in the Withaninae are discussed in the following. Finally, the results are embedded in an evolutionary context to explain the emergence of *MPF2-like* genes with respect to ICS formation in Physaleae.

4.1 The molecular basis of inflated calyx syndrome (ICS)

Morphological innovations are the hallmark of complexity and diversity of life. The appearance of these novelties in different body parts of plants provides them with possibilities to exploit the ecological niches, which eventually lead to an organismal adaptability (Mondragon-Palomono and Theissen, 2008). Morphological novelties in evolution usually arise due to changes in the developmental control genes (Theissen, 2000). The inflated calyx syndrome exhibited by Withania and Physalis is one of the examples of novel floral morphologies. In Withania the calyx undergoes a striking change in architecture during flower and fruit development, ultimately forming a balloon-like structure that encloses the mature berry.

The inflated calyx syndrome (ICS) formation in Physalis requires the expression of a MADS-box transcription factor *MPF2* - an ortholog of Solanum *STMADS16* gene, expressed strictly in the vegetative tissues - in the floral organs, especially in the calyx (He and Saedler, 2005). But this is not the only component involved in ICS; action of some plant hormones such as cytokinin and gibberellin is also essential (He and Saedler, 2005; He and Saedler, 2007). In Withania, it has been shown that crosstalk occurs between the developing berry and sepals. If sepals are removed at early stages in development, no fruit is elicited. On the other hand, if fertilization is prevented, no further growth of the calyx is observed. The growing pollen tube,

the fertilization process itself, or the developing fruit may provide signals needed for ICS formation. The cytokinin (6-BAP, 2-IP, zeatin) and gibberellin (GA3) are implicated as signals that promote ICS formation in Withania. Furthermore, neither mock treatment with water nor the addition of auxins could restore the calyx syndrome (He and Saedler, 2007). Therefore, ICS formation in Physalis requires 3 components in the calyx: an *MPF2-like* protein, cytokinin, and gibberellin (He and Saedler, 2007). Previously, the cytokinin 6-BAP has been shown to facilitate transport of MPF2 into the nucleus, thus promoting sepal cell division, and the resulting small cells elongate in response to gibberellin. The mechanism similar to ICS formation in Physalis intuitively, can be envisioned to be operating in Withania as well however, the interplay between hormones and *MPF2-like* genes in Withania needs to be tested empirically.

In addition to ICS formation, *MPF2* codes for MADS-box transcription factor otherwise involved in leaf formation and male fertility. This implies that all the three functions i.e. ICS, leaf development and male fertility are carried out by a unique gene *MPF2* in Physalis. In the present study, however, two classes of *MPF2-like* genes, *MPF2-like-A* and *MPF2-like-B* were isolated from Withania but only one class from Tubocapsicum, *MPF2-like-B*, since *MPF2-like-A* gene was shown to be lost from the genome. They are substantially different in their C-terminal region. The hallmark of MPF2-like-A proteins is a deletion of 3 amino acids in the C-domain and extension of an 8 amino acids at the C-terminus. Structurally, MPF2-like-B is the closed homolog of MPF2-like but phylogenetic analysis suggested that MPF2-like-A is more closely related to MPF2. Given the structural pattern differences, it is tempting to suggest the functional divergence between these two classes of MPF2-like proteins.

Therefore, isolation of two structurally different groups of MPF2-like genes suggests the occurrence of duplication in these genera.

4.2 Duplication of *MPF2-like* genes

Gene and genome duplications are prominent in plants and have greatly influenced their organization and evolution (Otto and Whitton, 2000; Wendel, 2000; Simillion et al., 2002; De Bodt et al., 2005; Maere et al., 2005). Duplications of genes are thought to provide the raw materials for the evolution of morphological novelties (Force et al., 1999; Carroll et al., 2001; True and Carroll, 2002; He et al., 2004; Harrison et al., 2005; Mondragon-Palomono and Theissen, 2008; Des Marias and Rausher, 2008). MADS-box proteins are integral components of plant development and often specify organ identity (Sommer et al., 1990; Schwarz-Sommer et al., 1990; Yanofsky et al., 1990; Theissen et al., 2000). MADS-box gene family

has greatly expanded in plants by a range of duplications that have enabled the genes to diversify in structure and function (Rijpkema et al., 2007). The number of these genes has increased considerably during the evolution of land plants, from roughly 20 in the moss *Physcomitrella patens* (Rensing et al., 2007) to more than 100 in higher eukaryotes like *Arabidopsis thaliana* (de Folter and Angenent, 2006; Parenicova et al., 2003; Rijpkema et al., 2007). Concomitantly plant architecture has also increased in complexity (Rijpkema et al., 2007).

While duplication of individual genes is one possibility to provide a playground for evolution, polyploidization, whether auto- or allo-polyploidization, is an alternative, which apparently has been taken in the Withaninae. The chromosome counts as well as the molecular studies suggested polyploidy in Withania and Tubocapsicum. Since duplicate sets of diverse genes were found for *MPF1-like* and *MPF2-like* genes as revealed by neighbour-net network analysis, allo-tetraploidization is the most likely type of ploidy in the two genera, Withania and Tubocapsicum. Usually paralogues retain the ancestral function shortly after duplication. As the evolution progresses, mutations render one of the two copies non-functional for there is no selective pressure to keep both the copies functional. The loss or non-functionalization (pseudogenization) is therefore the ultimate fate of the duplicated genes. However, a large number of duplicated genes escape their extinction, mainly because of sub-functionalization or neo-functionalization (Force et al., 1999; Prince and Pickett, 2002; Mondragon-Palomono and Theissen, 2008). Therefore, sub-functionalization and/or neo-functionalization can be the expected fate of *MPF2-like* genes after duplication.

4.3 Sub-functionalization and/or neo-functionalization of *MPF2-like* genes

The MADS-box proteins seem to have multiple functions (Riechmann and Meyerowitz, 1997) and duplication of the genes might allow evolution of novel functions and/or sub-functionalization (Force et al., 1999; Carroll et al., 2001; True and Carroll, 2002; He et al., 2004; Harrison et al., 2005; Irish and Litt, 2005). A prominent case of sub-functionalization, for example, concerns the duplication of *PLENA* (*PLE*) of *Antirrhinum majus* and its duplicate *FARINELLI (FAR)* (Davies et al., 1999; Causier et al., 2005). *PLE* determines organ identity of whorl 3 and whorl 4, i.e. a *ple* mutant features homeotic conversion of stamen to petals and carpels to sepals (Bradley et al., 1993), the flowers of *far* mutants are morphological normal but partially male sterile (Davies et al., 1999). In Arabidopsis *agamous* mutants have a similar phenotype as *ple* mutants in Antirrhinum, but its duplicates, the redundant *SHP-1* and *SHP-2* (Irish and Litt, 2005) control seed dispersal in Arabidopsis

(Liljegren et al., 2000). Another example of diversification through sub-functionalization is rice *AGAMOUS* (*AG*)- clade genes *OsMADS3* and *OsMADS58*. The *AGAMOUS* in Arabidopsis regulates reproductive organ identity and floral meristem. But *OsMADS3* and *OsMADS58* have divided the functions. The *OsMADS3* is mainly recruited for floral meristem specification and carpel morphogenesis whereas *OSMADS3* specify stamen identity and inhibit lodicules development (Ymaguchi et al., 2006).

The situation of sub-functionalization is obscure in Petunia's orthologs *PMADS3* and *FBP6* (Tsuchimoto et al., 1993; Kater et al., 1998), the orthologs of the morning glory *DP* and *PN* and the orthologs *C1* and *D1* of roses (Irish and Litt, 2005).

Until now, no discrete examples of studies concerning gene duplication followed by positive selection and neo-functionalization in plants are available except some enzyme encoding genes such as methylthioalkylmalate synthases (*MAM*) involved in glucosinolate pathway and dihydroflavonol-4-reductases (*DFR*) encoding enzymes involved in anthocynin pathway (Benderoth et al., 2006; Des Marais and Rausher, 2008). In most cases the results are based on only computational methods and lack the complete functional data. An elegant example of neo-functionalization is the inflated calyx syndrome (ICS) in Physalis, which requires the heterotopic expression of a MADS- box transcription factor *MPF2*. The absences of two CArG motifs in the upstream promoter region of *MPF2* gene are accounted for floral expression of this gene in Physalis, which eventually lead to ICS formation in this genus (He and Saedler, 2005). As previously mentioned all the three functions i.e. ICS, leaf development and male fertility are carried out by a unique gene *MPF2* in Physalis. However, the two *MPF2-like* genes i.e. *MPF2-like-A* and *MPF2-like-B*, duplicated due to allopolyploidy in Withania species differ structurally and they seem to have undergone mutational changes prior to their evolution in the Withaninae. At the onset of evolution each of the two genes are assumed to have possessed all or some of three functions in leaf and calyx development and in male fertility. Upon polyploidization, they might have undergone further mutations in gene or *cis*-regulatory elements affecting the expression pattern of the genes and the protein structure, thus leading to sub-functionalization and/or neo-functionalization (Fig. 4.1).

Over-expression and yeast two-hybrid analyses may provide the functional evidences for possible occurrence of sub-functionalization and/or neo-functionalization of *MPF2-like* genes.

4.3.1 *MPF2-like-A* genes are involved in the formation of enlarged sepals in transgenic Arabidopsis

Previously it was shown that *MPF2*-gene expression is plesiomorphic and hence it was assumed that ICS formation might be plesiomorphic as well, at least in the Physaleae (Hu and Saedler, 2007). However, some Witheringia and Vassobia species expressing the *MPF2-like* gene in floral tissues did not form ICS. Secondary mutations might have generated this diversity, but only in *Witheringia coccoloboides* the site in the ICS pathway could roughly be identified (Hu and Saedler, 2007).

In *Tubocapsicum anomalum MPF2-like-A* is missing and *MPF2-like-B* is expressed in floral buds, but nonetheless no ICS is formed. Since, converse to *MPF2* (He and Saedler, 2006) or *MPF2-like-A,* over-expression of *MPF2-like-B* in transgenic Arabidopsis did not yield large sepals suggesting that *MPF2-like-B* is deficient with respect to ICS formation in Tubocapsicum as well and thus supports the hypothesis of secondary loss of ICS formation in the Physaleae.

The observation that *MPF2-like-A* rather than *MPF2-like-B* is expressed in floral tissues in *Withania somnifera,* might suggest that *MPF2-like-A* controls ICS formation in this species. This assumption is corroborated by the large sepals phenotype of transgenic Arabidopsis plants over-expressing *MPF2-like-A* (*WSA206*). The enlarged sepals phenotype mimics the *AGL24* over expressor (He et al., 2004). But some differences between the two over-expressors i.e. *WSA206* and *AGL24* however, do exist. Firstly, *AGL24* over-expressor, in most cases harbours a secondary flower without petals (Liu et al., 2007). Secondly, the pedicel length of *WSA206* over-expressor is tremendously increased. Interestingly, phenotype does not require external application of cytokinin in contrast to *MPF2* and *STMADS16*, which essentially need the external administration of cytokinin to generate lantern-like structure in Arabidopsis reminiscent of ICS in Physalis (He and Saedler, 2007). But this enlarged sepal growth without cytokinin is coincident with foliose sepal formation in *A. thaliana* as a result of ectopic expression of *ZMM19* an ortholog of *MPF2* in *Zea mays* (Münster et al., 2002a).

Generally, the transgenic plants transformed with YFP gene driven by CaMV 35S promoter are envisioned to show strong YFP expression ubiquitously in all tissues and cells. In this experiment, though strong YFP signals were detected in the sepals of *35S:WSA206* plants but signals from plants transformed with *WSB206, TAB201* genes and empty vector were weak. Low intensity of YFP signals in transgenic plants may be due to interference by defence mechanism of the host plant. There are proteins known which can suppress degradation by host plant. The P19 is one of a class of plant and animal virus proteins that are known to

interfere with a host defence mechanism, referred to as RNA silencing, that targets foreign (virus and retrotransposon) RNAs in a sequence-specific manner for degradation (Molnar et al., 2005). Therefore, a transient infiltration of *N. benthamiana* leaves with constructs containing MPF2-like proteins mixed with P19 proteins is proposed to overcome the host defence and to improve the YFP signal strength.

The phenotype of *MPF2-like-B* (*WSA206*) over-expressing transgenic Arabidopsis is indistinguishable from wild type Arabidopsis plants suggesting that this gene is also deficient in ICS formation similar to *MPF2-like-B* of Tubocapsicum.

Therefore, the functions of MPF2-like-A and MPF2-like-B differ. Unequivocally *MPF2-like-A* genes are involved in ICS. Apart from their putative role in leaf development the function of *MPF2-like-B* genes remains obscure.

4.3.2 Arabidopsis floral developmental control proteins interact with MPF2-like-A but not with MPF2-like-B proteins

MPF2 is known to interact with other MADS-box proteins to form hetero- or even higher oligomers (He et al., 2007). These might control cell division either in leaves or the calyx depending on their interacting partners.

The differentiation of MPF2-like proteins is further evident from their different affinities for Solanaceous and Arabidopsis interactors. The interactions of MPF2-like-A (WSA206) proteins preferentially with Arabidopsis proteins such as SEP1, SEP2, SEP3, SEP4, SOC1, MAF1, MAF3, MAF4 etc. involved in regulating flowering, supports the notion that these proteins might be involved in ICS formation via interacting with flowering regulatory proteins. The inability of MPF2-like-B (WSB206) proteins to interact with any of the Arabidopsis flowering proteins suggests the different functions for them, other than the ICS. Their interactions with IAA16, EMB1417, ATSIP, AKIN and MYB4 proteins might allude towards the role in leaf, embryo, seed and/or root development. With Tubocapsicum MPF2-like-B protein (TAB201), the interaction of a few flowering proteins (SOC1, MAF1 and MAF5) is correlated with its native expression in the flower. The overall interaction patterns suggest that Tubocapsicum MPF2-like-B proteins are intermediate between Withania MPF2-like-A and MPF2-like-B proteins of Withania. On one hand they are interacting with Arabidopsis proteins involved in regulating flower development like Withania MPF2-like-A (WSA206), whereas on the other hand, they interact with non-MADS-domain proteins similar to Withania MPF2-like-B (WSB206) proteins. Hence, Tubocapsicum MPF2-like-B proteins have a mixed behaviour. The interaction of all the MPF2-like proteins with non-MADS-

domain proteins (AKIN10, C2 domain containing protein, PADI, EMBRYO DEFECTIVE, unknown Arabidopsis proteins) reflects at least some common functions among them, which may likely be the leaf development.

4.3.3 Solanaceous MAGO NASHI-like proteins interact with MPF2-like-B proteins but not with MPF2-like-A

One of the important functions of MPF2 in Physalis is the male fertility function. Since, however, MPF2 interacts also with PFMAGO, an ortholog of mago nashi from Drosophila known to control formation of progeny (Boswell et al., 1991; Newmark and Boswell, 1994), a role it might serve in Physalis as well (He et al., 2007).

The first mago nashi gene (meaning "no grandchildren" in Japanese) was identified in Drosophila, as a strict maternal effect gene. It is required for the formation of the embryonic axes and for germ-cell determination (Boswell et al., 1991; Newmark and Boswell, 1994; Micklem et al., 1997; Newmark et al., 1997). It always functions together with an RNA-binding protein, known as Y14 in Xenopus (Kataoka et al., 2000), RBM8A in humans (Zhao et al., 2000), and Tsunagi in Drosophila (Mohr et al., 2001), a protein that shuttles between nucleus and cytoplasm (Hachet and Ephrussi 2001, 2004; Kim et al., 2001). The 1st mago nashi homolog found in plants was isolated and characterized from *Oryza sativa* (Swidzinski et al., 2001) and mago nashi mutations were found to cause sterility in *A. thaliana* (Johnson et al., 2004; Pagnussat et al., 2005).

While investigating fertility function, scenarios similar to MPF2 can be assumed for MPF2-like proteins as well. When both MPF2-like-A (WSA206) and MPF2-like-B (WSB206) proteins from Withania are tested for their interaction with the MAGO NASHI-like proteins from Withania (WSMAGO1 and 2-), Tubocapsicum (TAMAGO1 and 2-) and Physalis (PMAGO1 and 2-), surprisingly only MPF2-like-B proteins could show the affinities for the Solanaceous MAGO NASHI-like proteins. This is the first hint for the function of MPF2-like-B proteins. It leads to the assumption that there is a possible role of MPF2-like-B protein in fertility because MPF2 together with PFMAGO1 and 2- is probably, involved in fertility function in Physalis. An expression study of *MAGO NASHI-like* genes in Withania and Tubocapsicum might reveal their role in fertility pathway. The ability of Tubocapsicum MPF2-like-B proteins to show affinities for some of the MAGO NASHI-like proteins further support their intermediate role between Withania MPF2-like proteins.

4.3.4 Arabidopsis flowering time control proteins interact with MPF2-like-A proteins but not with MPF2-like-B

The MADS-domain genes are involved in the transition to flowering in Arabidopsis (Borner et al., 2000; Masiero et al., 2004; Lee et al., 2007; Kim et al., 2008; Veley et al., 2008). The svp mutant shows early flowering without displaying other obvious phenotypes (Hartmann et al., 2000). Loss or reduction of AGL24 activity by double-stranded RNA (dsRNA)-mediated interference results in dosage-dependent late flowering, while over-expression of *AGL24* under the CaMV 35S promoter leads to early flowering (Yu et al., 2002b). Suppression of flowering time genes like *SVP*, *AGL24* and *SOC1* by *AP1* or *LEAFY* results in the transition from vegetative to floral phase (Liu et al., 2007). The flowering time control genes *AGL24* and *SVP* belonging to *STMADS11* gene clade (Gregis et al., 2006, Liu et al., 2007) are the so-called orthologs of *MFP2-like* genes in Arabidopsis. Therefore, in addition to their role in ICS, fertility and leaf development, a further role of MPF2-like-A and Tubocapsicum MPF2-like-B proteins in controlling flowering time can be assumed intuitively. This assumption is supported not only by their interactions with AGL20 and MAF1, MAF3 and MAF4 proteins but also by conservation of many sites for INDETERMINATE1 (ID1) in all the *MPF2-like* promoter and first intron sequences. Hence, Withania MPF2-like-A (WSA206) and Tubocapsicum MPF2-like-B (TAB201) proteins may be recruited in regulating the flowering time, unlike WSB206 protein which shows no affinity for proteins involved in flowering time control, in yeast system. At present, the experiment concerning the role of MPF2-like proteins in flowering time control in transgenic Arabidopsis is undertaken.

The rationale for differences in the interaction properties of MPF2-like proteins may be due to their highly variable C-terminal region, which is usually known to influence the interacting properties. This is supported by the presence of an 8 aa long repressor motif (R-motif) at the C-terminus of MPF2-like-A proteins. Removal of R-motif from MPF2-like-A proteins and its addition to MPF2-like-B proteins brings about the self-activation of both the proteins. At present domain swapping experiments in yeast two-hybrid system are underway to characterize the role of R-motif fully.

Therefore, MPF2-like-B might have different functions. This might already be reflected in their expression pattern.

4.3.5 Expression of *MPF2-like-A* and *MPF2-like-B* genes differs in the two genera

MPF2-like-A (*WSA206*) is expressed in Withania species in vegetative as well as in floral tissues; expression of the *MPF2-like-B* (*WSB206*) is restricted to vegetative tissues. This is

different in Tubocapsicum in which *MPF2-like-B* (*TAB201*) is also expressed in floral organs. Promoter sequence analyses revealed that a highly conserved CArG-box is missing in Withania's *MPF2-like-B* gene up-stream of transcription initiation site. Some MADS-box proteins involved in flower development recognize this DNA-binding motif (de Folter and Angenent, 2006). The presence of this motif in the promoter of *MPF2-like-B* of Tubocapsicum might assure floral expression and hence might contribute to fertility of this species, as does MPF2 in Physalis (He and Saedler, 2005). This is corroborated by interaction of TBA201 with SEP1, SEP3, SEP4, SOC1, MAF1 and MAF4, which are flowering proteins expressed in flower. Besides, Tubocapsicum MPF2-like-B (TAB201) proteins exhibit affinities for Solanaceous MAGO NASHI-like proteins, the orthologs of mago nashi from Drosophila known to control formation of progeny (Boswell et al., 1991; Newmark and Boswell, 1994), assumed to be involved in fertility. However, only complementation of an *MPF2-like* mutant deficient in male fertility could verify the speculation that TAB201 is needed for male fertility in Tubocapsicum. Unfortunately no such mutant is currently available.

However, in Withania *MPF2-like-B* is not expressed in floral organs and the conserved CArG-box is missing in its promoter. Since expression of MPF2-like-B (WSB206) form Withania may be masked by MPF2-like-A (WSA206) in the leaf tissues, a question concerning the function MPF2-like-B still remains to be answered. Although a slight hint for the role of MPF2-like-B in Withania is available in the form of its interaction with Solanaceous MAGO NASHI-like proteins. But this protein does not interact with any of flowering proteins interactors such as SOC1, SEP1, SEP3, SEP4, MAF1, MAF3 and MAF5 and natively, is not expressed in the flower, which rules out its possible role in fertility. In these scenarios, the male fertility function in this polyploid species thus might be provided by MPF2-like-A. This is supported by phenotypic alterations in the silique morphology while over-expressing this gene in Arabidopsis. In severe cases the siliques are devoid of any seeds. Furthermore, interactions of MADS-domain flowering proteins such as SOC1, SEP1, SEP2, SEP3, SEP4, MAF1, MAF3 and MAF4 exclusively with MPF2-like-A substantiate its possible role in fertility. A few questions remain elusive: why is *MPF2-like-B* from Withania not lost from its genome, as is *MPF2-like-A* from Tubocapsicum, is it still under selection and if so for which function of MPF2-like?

4.3.6 *MPF2-like* genes are differently regulated

Although expression of *MPF2-like-B* genes in Withania is confined to leaves only, but promoter GUS expression analyses in Arabidopsis suggested a strong activation of GUS gene in inflorescence driven under *MPF2-like-B* promoters. The floral expression is unexpected in the case of *MPF2-like-B* genes of Withania. Many speculations can be put forward to explain this paradoxical expression. One possibility, which can affect expression, is the heterologus nature of the host. Arabidopsis is evolutionary fairly distant to Withania. He (unpublished results) experienced similar results while analyzing the promoter GUS expression for *MPF2* and *STMADS16* genes in transgenic Arabidopsis. Natively, *STMADS16* is expressed in the leaf tissue only and *MPF2* both in the flower and leaves. Contrary to expectations, in transgenic Arabidopsis the expression patterns were reversed. *STMADS16* expression in comparison with *MPF2* was stronger in flowers. Another possibility can be that Withania *MPF2-like-B* genes are expressed in flower tissue but very weakly. A detailed expression study of Withania *MPF2-like* genes in individual floral organs with Real Time PCR and *in situ* hybridization can help to uncover the expression differences between *MPF2-like-A* and *MPF2-like-B* genes completely.

Differences in the regulation of *MPF2-like-A* and *MPF2-like-B* genes are also evident from their different intron-GUS expression patterns in transgenic Arabidopsis. Generally first intron suppresses the expression of *MPF2-like* genes in leaves. With Mulan analysis, in addition to promoter region, two CArG-boxes were identified in each of the 1st intron of *MPF2-like* genes. Evidence is growing on the presence of regulatory elements in introns. Introns can significantly affect gene expression in plants and many other eukaryotes in a variety of ways. Some introns contain enhancer elements or alternative promoters. Both positive and negative regulation can be mediated by regulatory sequences in the introns (Rose, 2004, 2008). Liu et al., (2007) identified a CArG-box in the third intron of *AGL24* gene of Arabidopsis, which is a putative AP1 binding site. Mutation of this CArG-box causes the ectopic expression of *AGL24* in young floral meristem.

In this experiment the results demonstrate that no GUS signals are detected in Arabidopsis leaf tissues when transformed with promoter plus intron of *MPF2-like-B* genes. Similarly, very weak YFP signals are detected in *N. benthamiana* leaves transiently transformed with promoter plus intron. On the other hand strong signals for GUS and YFP are observed in leaves, transformed with promoter only. These observations suggest that introns are suppressing the expression of *MPF2-like* genes. The potentials explanations for the suppression of GUS and YFP expression in leaf tissue by introns are many. Firstly, the length

of the insert is around 5kb, which can cause problems in translation. Secondly, inclusion of MADS-domain in the insert may translate a truncated protein, which may repress the GUS gene expression. Thirdly, the presence of some repressor elements in the introns may have inhibitory effect on the expression. The third possibility seems to be a more plausible reason for lack of GUS expression as well as the low intensity of YFP signals. Because while analyzing intron sequences, AuxREs (auxin response factor elements) which are binding sites for ARF (Auxin response factor) could be identified in all the intron sequences except *WSA206* (*MPF2-like-A*). The AuxREs can mediate both activation and repression (Ulmasov et al., 1999a; Ulmasov et al., 1999b; Guilfoyle and Hagen, 2007). Auxin response factor elements play a role in the regulation of plant growth and development (Lisum, 2002). In this experiment GUS expression in Arabidopsis leaves is exhibited by the intron-GUS construct of this gene only while all the others are suppressing the expression. Hence ARF can regulate the expression of *MPF2-like* genes in Arabidopsis, which in most cases is suppression.

4.4 Different models for the expected fate of *MPF2-like* genes

It is already mentioned that MPF2-like proteins have duplicated in the Withaninae probably, due to allo-tetraploidization. A number of models for the expected fates of duplicated genes have been proposed previously. A few of these include neo-functionalization (NEO-F), duplication-degeneration-complementation (DDC) and escape from adaptive conflict (EAC) (Mondragon-Palomono and Theissen, 2008; Des Marais and Rausher, 2008). Normally, non-functionalization is a frequent fate of duplicated genes. However, most duplicated genes escape their extinction because of sub-functionalization or neo-functionalization (Mondragon-Palomono and Theissen, 2008). In case of sub-functionalization the original function of a gene is partitioned between its descendants mainly because of mutations in the *cis*-regulatory elements in each paralogue (Force et al., 1999; Prince and Pickett, 2002; Mondragon-Palomono and Theissen, 2008). If this consideration is applied to Withania *MPF2-like-A* and *MPF2-like-B* genes, both share the function of leaf development with alterations in the promoter sequences such as, the presence of CArG-box in *MPF2-like-A* and absence in *MPF2-like-B*. So basically, this fits to the occurrence of sub-functionalization in these genes. However, marked differences in structures of *MPF2-like-A* genes especially, a 3 aa deletion and an 8 aa extension at the C-terminal region and positive Darwinian selection suggests a new function for this gene. According to NEO-F model, novel functions arise after gene duplication, with one copy maintaining ancestral function whereas the second copy undergoes an adaptive change and selected to perform a new function. If only Withania *MPF2-like-A*

and *MPF2-like-B* genes are concerned, MPF2-like-B maintains the basic function of leaf development while MPF2-like-A after positive selection, along with changes in *cis*-regulatory elements in the promoter regions, gain a new function in the form of expression of MPF2-like-A in the flower and eventually control of ICS formation, lead to nothing but neo-functionalization of this gene. According to EAC model the adaptive mutation occurs in both the duplicated copies. This, intuitively, is not the scenario here because MPF2-like-A is selected positively whereas MPF-like-B is under purifying selection. In conclusion, none of the models is really applicable to ICS evolution when more species samples are included. The situation become complicated when Tubocapsicum is included which has lost *MPF2-like-A* genes and *MPF2-like-B* is impaired in ICS formation.

MPF1-like genes constitute another group of genes, which are duplicated as a result of allo-tetraploidization in the Withninae. There is no gene loss in the case of Tubocapsicum since both *MPF1-like-A* and *MPF1-like-B* could be isolated. In comparison to MPF2-like proteins, MPF1-like proteins are not divergent in their structures as dramatically as MPF2-like and show no positive Darwinian selection during directional evolution. Hence occurrence of neo-functionalization is naturally not expected from these genes in the Withaninae.

4.5 Evolution of *MPF2-like* genes and origin of ICS

As already mentioned in the previous section that expression of *MPF2-like* genes in the Physaleae is a plesiomorphic trait and hence the development of ICS might be a basic trait as well (Hu and Saedler, 2007). According to this assumption all the Physaleae naturally should form ICS. However, in some species of the Physaleae such as *Witheringia solanacea, Witheringia coccoloboides* and *Vassobia breviflora* even though *MPF2-like* gene expression is observed no ICS is formed. During evolution of these species they might have lost ICS, due to secondary mutations in the ICS pathway (He and Saedler, 2007; Hu and Saedler, 2007).

However, plesiomorphy of gene expression does not exclude an origin of ICS. In the case of Withania and Physalis two independent events could have led to ICS during evolution of these genera. The two protein types MPF2-like-A (Withania) and MPF2-like (Physalis) have quite divergent signatures: a 3 amino acid deletion and an 8 amino acid long extension at the carboxyl-terminal end characterizes MPF2-like-A, both features are missing in MPF2-like proteins. In the evolution of ICS in Withania and in Physalis both proteins were underlying positive Darwinian selection prior to separation of the genera and speciation within the genera. These latter branches were under purifying selection resulting in evolutionary stability of these proteins and thus maintaining the newly evolved trait: ICS. However, 3 aa deletion

and 8 aa extensions patterns suggests that this scenario is also unlikely to prevail, since such features could have arisen in response to visitations of transposon during evolution of these proteins. Transposons could have generated an MPF2-like-A protein from an MPF2-like progenitor. Plant transposons are known to generate deletions and insertions of DNA sequences upon excision from their site of integration (Schwarz-Sommer et al., 1985). If the site of visitation happened to occur within an exon, a frameshift mutation could result from excision of the element (Paz-Ares et al., 1990; Singer et al., 1998). C-terminal extensions of MADS-box proteins due to frameshift mutations have been discussed in relation to sub-(neo)-functionalization of such proteins (Vandenbussche et al., 2003).

The three amino acid deletion is a signature of MPF2-like-A and is shared by STMADS16 of *S. tuberosum,* Nt002 of *Normania triphylla*, Ss001 of *Solanum sisymbriifolium* and the MSM2 of *S. macrocarpon* (He and Saedler, 2005). However, the 3 amino acid deletion seems not to contribute to ICS formation, mainly because *S. sisymbriifolium* and *S. macrocarpon* feature large sepals upon pollination (data not shown). This is supported by the ICS-like phenotype of transgenic *S. tuberosum* plants over expressing STMADS16 in its native host provided the plant hormone cytokinin (6-BAP) is also available. Moreover, the large sepals phenotype of transgenic Arabidopsis plants over expressing STMADS16 in the presence of cytokinin further strengthen this assumption (He and Saedler, 2006).

On the other hand, the effect of the 8 amino acid long extensions another hallmark of this protein of MPF2-like-A proteins on ICS function remains elusive. Domain swapping experiments followed by phenotypic studies of transgenic plants are under way and might clarify the situation.

In such a scenario, the common progenitor of MPF2-like-A of Withania and MPF2 of Physalis had already ICS function and therefore selection of ICS would have been a monophyletic event. However, currently the distribution pattern and the functions of both types of genes within the Solanaceae are unfortunately not known. A broader sampling within the basal Physaleae including Cuatresia (Olmstead and Bohs, 2007) might be revealing in the assessment of the proposed origin of ICS. In any case, once the two genes *MPF2-like-A* and *MPF2-like-B* had evolved in independent lineages sub- (neo)-functionalization could have commenced in a polyploid lineage, like in the Withaninae.

On first glance MPF2-like-B is more similar in protein patterns to MPF2-like than is MPF2-like-A, but functionally it has lost the capacity to cause ICS. However, despite their sequence disparity the latter two proteins share the function of ICS formation, possibly due to similar selection pressures.

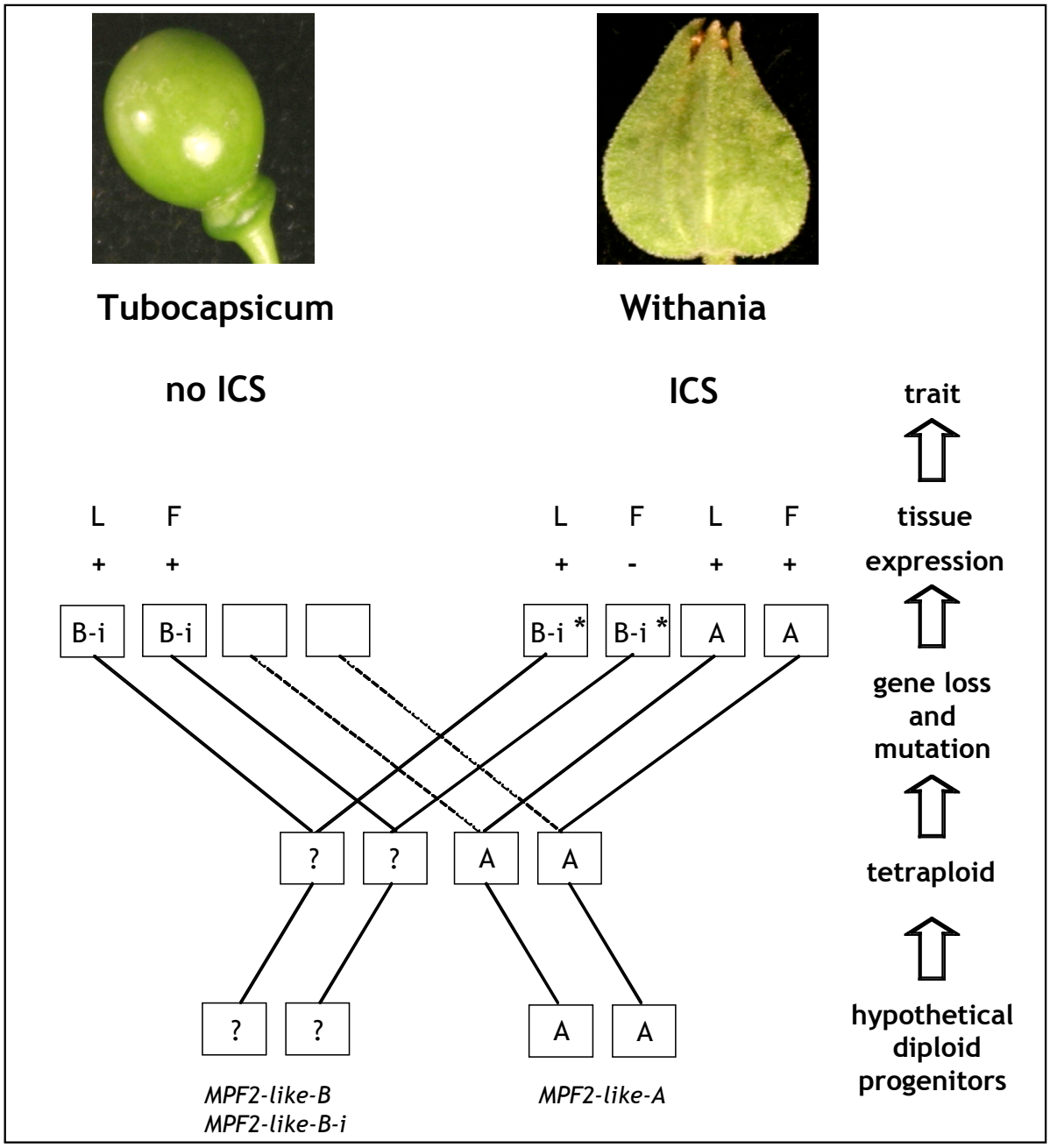


Figure 4.1: Evolutionary events of *MPF2-like* genes history

Based on structural and functional characteristics of *MPF2-like* genes, the proposed evolutionary history of these genes is shown. It is presumed that *MPF2-like* genes had already diverged in their functions. Their hypothetical diploid progenitor somehow, was proficient in carrying out some or all the functions of *MPF2-like* genes in the tetraploid Withania and Tubocapsicum. These functions include leaf development, ICS formation and male fertility. During evolution two possibilities could exist. One is that mutations might have accumulated in these two groups of *MPF2-like-A* and *MPF2-like-B* genes prior to allo-tetraploidization. Alternatively, several mutational events might have occurred post allo-tertaploidization in the two lineages. During this period both the lineages followed normal speciation procedures and acquired their evolutionary fates i.e. no functional change, pseudogenization, gene loss, promoter innovation, sub- and/or neo-functionalization etc. In the Tubocapsicum

lineage, *MPF2-like-A* gene was lost from the genome. While mutations rendered *MPF2-like-B* gene deficient in ICS formation although *MPF2-like-B* gene is certainly expressed in floral organs as well, but clearly no ICS is observed in reality. Mutant *MPF-like-B* sequence is denoted by B-i while wild type is shown as B. Contrary to Tubocapsicum, *MPF2-like-A* gene in Withania somehow carried out its normal function in the form of ICS formation. But *MPF2-like-B* went under a promoter mutation such that its expression became restricted to the leaf tissue only and hence became impaired in ICS formation. An asterisk marks promoter alteration of *MPF2-like-B* gene in Withania.

4.6 Different selection pressures and diversity of ICS in Withania species

Rapid evolution driven by positive Darwinian selection is a recurrent theme in adaptive evolution; however, positive selection has never been demonstrated for proteins involved in morphological novelties. Therefore, directional evolution of MPF2-like-A and MPF2-like-B proteins was tested in the Withaninae and MPF2-like proteins in Physalis. In the evolution of ICS in Physalis and in Withania a progenitor of both proteins seem to have been under positive Darwinian selection prior to separation of the genera and speciation within the genera. Later in evolution, however, purifying selection apparently dominated resulting in evolutionary stability of the proteins and thus maintaining the newly evolved trait: ICS. Therefore, Darwinian selection of *MPF2-like-A* protein in Withania might indicate selection of the ICS trait itself. Speculations concerning the selection pressure to generate or maintain ICS are many folds. One possibility involves generation or maintenance of a microclimate within the lantern providing humidity for the developing berry in a dryer environment. This would lead to the hypothesis that, the higher the humidity, the lower the degree of calyx inflation. For this purpose, average rainfall in different regions of the world, to which Withania and Tubocapsicum are endemic, was recorded from different meteorological sources on the internet (data not shown). The data seem to support the above hypothesis. The genus Tubocapsicum prefer humid climate. As the developing berries do not need humidity so it has only a rudimentary calyx. Merrill (1923) found the non-ICS forming Tubocapsicum exposing its berry in humid environments, i.e. growing in “mossy forest” at 1800 m to 2200 m and along streams in shaded ravines at low and medium altitudes in the Philippines. Withania species displaying semi-succulent and papery ICS, especially *W. somnifera* and *W. coagulans* perfectly covering their berries prefer dryer habitats (Hepper, 1991) for example in north western India and in the Baluchistan (average rainfall is 50-70 mm per annum) province of Pakistan (Negi et al., 2000). The fleshy berries of *W. aristata* endemic to xeric parts of the Canary Islands are part of the frugivorous diet of the lizard *Gallotia galotti* (Valido et al., 2003), which might indicate that the lantern of this plant maintains the juiciness of the berry

throughout the dry season. So not only the *MPF2-like-A* gene but also the ICS trait itself may be selected as revealed by the habitat preference by different species of the genera.

In conclusion, gain and loss of function played a role in the evolutionary history of the Physaleae as exemplified by positive Darwinian selection of MPF2-like proteins leading to ICS formation and secondary mutations affecting this pathway and modifying the trait (He and Saedler, 2007; Hu and Saedler, 2007). This might be an ongoing process, which can be studied further in the basal Physaleae.

In a nutshell, the methodology of earlier studies concerning the duplications and diversification of genes involved in evolutionary novelties relied largely on DNA sequence analysis but lacked functional assays of duplicated genes, frequently generating ambiguous results. Here in this study an amalgamation of both computational and experimental approaches has been used to address the questions related to duplication, sub- (neo)-fictionalization, selection and diversification of genes underlying the evolutionary morphological novelties. These findings demonstrate the contribution of gene duplication to organismal adaptability and show the power of combining sequence analysis and functional assays in delineating the molecular basis of morphological novelties. The findings reported here therefore, are not only of fundamental significance in explaining how the calyx novelty evolved was maintained in Physaleae. But also provide a nice example of using wild 'evolutionary mutants' to dissect complex trait variations. Moreover, these findings will also promote the study of Darwinian selection in understanding and maintaining the biodiversity.

5 Abstract

This thesis concerns the molecular basis of the evolution of morphological novelties in Solanaceae. The aim is to contribute to understanding the sub-functionalization and/or neo-functionalization of genes involved. The methodology of earlier studies relied largely on DNA sequence analysis but lacked functional analyses of duplicated genes, frequently generating ambiguous results. In this study an amalgamation of both computational and experimental approaches has been employed to address the questions related to duplication, sub- and/or neo-functionalization, selection and diversification of genes underlying the evolutionary morphological novelty of the Inflated Calyx Syndrome (ICS) prevailing in the Withaninae. In the genus Withania, like in Physalis, the sepals resume growth after pollination to encapsulate the mature fruit, forming a balloon-like structure termed ICS. While Withania species display a diversity of inflated calyces, Tubocapsicum, a close relative, has only a rudimentary calyx. Plant hormones such as cytokinin and gibberellin can mimic the fertilization signals and are essential for ICS formation in Withania. In the diploid species *Physalis floridana*, a unique gene *MPF2* belonging to the MADS-box gene family, in addition to controlling ICS formation, is responsible for leaf development as well as male fertility. However, Withania being a tetraploid, as evident from ploidy evaluation, offers the possibility to sort out these functions. Two classes of *MPF2-like* orthologs, *MPF2-like-A* (i.e. *WSA206*) and *MPF2-like-B* (i.e. *WSB206*), were isolated from Withania, but only one class from Tubocapsicum, *MPF2-like-B (TAB201)*, since an *MPF2-like-A* gene does not exist in the genome. Although their gene structures are similar, MPF2-like-A proteins have an aberrant structure with respect to MPF2 of Physalis. MPF2-like-A proteins feature a 3aa deletion in their C-terminal region and an 8 aa addition at their C-terminus. Nonetheless, they seem to be phylogenetically closer related to MPF2 than the MPF2-like-B proteins. Positive Darwinian selection was observed in the branch leading to Physalis MPF2-like and Withania MPF2-like-A proteins. MPF2-like proteins of non-ICS forming species as well as all MPF2-like-B proteins were under purifying selection. Over-expression of an *MPF2-like-B* gene (*WSB206*) in Arabidopsis did not demonstrate a phenotype deviating from wild type, but *MPF2-like-A* (*WSA206*) caused enormous enlargement of sepals and alterations in the number as well as the morphology of fruits indicating that the functions of the two proteins are different. This is supported by their divergent expression patterns. In Withania the *MPF2-like-A* gene is expressed in vegetative and in flower tissues and thus correlates with ICS formation, while the expression of *MPF2-like-B* gene is restricted to vegetative tissues only. This is different in

Tubocapsicum where the *MPF2-like-B* gene is expressed in all tissues tested but ICS formation is impaired.

Bioinformatic analyses of promoter and first intron sequences of *MPF2-like* genes have revealed that a conserved CArG-box - putative binding site for AG, SEP1, SEP4, AGL1, AGL15 - close to the transcription initiation site is present in all the *MPF2-like* promoters except *MPF2-like-B* of Withania (*WSB206*), while auxin response elements (AuxREs) are not present in *MPF2-like-A,* but found in all the *MPF2-like-B* genes in a 100 base pair highly conserved region at different positions in the first introns. Promoters of both the *MPF2-like-A* and *MPF2-like-B* genes drive strong GUS expression in transgenic Arabidopsis flowers, although *MPF2-like-B* (*WSB206*) is not expressed in the flower natively. Remarkably, the first introns generally suppress the activation of reporter gene (GUS and YFP) expression both in stable Arabidopsis transformants and transient assay of *N. benthamiana* leaves. Only MPF2-like-A proteins have the ability to form homodimers in yeast two-hybrid analysis. The repressor motif (R-motif) at the C-terminus of MPF2-like-A proteins is involved in the self-activation. Arabidopsis oligo-dT yeast library screening with these proteins yielded numerous non-MADS and MADS-domain interactors. The interactions of MPF2-like-A proteins with Arabidopsis proteins involved in regulating the flowering time (e.g. SOC1, AGL27, AGL19, AGL31, AGL42, AGL68, MAF4, SEP3 and SEP4), and of MPF2-like-B proteins of Withania exclusively with non-MADS-domain proteins (e.g. IAA16, EMB1417, ATSIP, AKIN and MYB4) support their functional divergence and coincide with their different expression patterns. Tubocapsicum MPF2-like-B (TAB201) proteins can interact with SOC1, AGL27, and MAF5 indicating their intermediate role between Withania MPF2-like (WSA206 and WSB206) proteins. However, Withania MPF2-like-B proteins can interact with MAGO NASHI homologs of Withania, Tubocapsicum and Physalis. The gene mago nashi is a highly conserved protein known to be involved in many developmental processes including germ cells differentiation. Withania MPF2-like-A cannot interact with Solanacaeous MAGO NASHI-like proteins (WSMAGO1, WSMAGO2, TAMAGO1, TAMAGO2, PFMAGO1, PFMAGO2) but the ability of Tubocapsicum MPF2-like-B proteins to show affinities for some of the MAGO NASHI-like proteins further support their intermediate role between Withania MPF2-like proteins.

Unequivocally *MPF2-like-A* genes are involved in ICS. Apart from their putative role in leaf development, the function of *MPF2-like-B* genes remains obscure. The contribution of allo-tetraploidization to sub- and/or neo-functionalization of *MPF2*-like genes are discussed, as is the positive Darwinian selection in the evolution of the Physaleae. These findings

demonstrate the contribution of gene duplication to organismal adaptability and show the power of combining sequence analyses and functional assays in delineating the molecular basis of morphological novelties. The findings reported here are therefore not only of fundamental significance in explaining how the calyx novelty evolved and was maintained in Physaleae, but also provide a nice example of using wild 'evolutionary mutants' to dissect complex trait variations.

6 Zusammenfassung

Diese Arbeit befasst sich mit der molekularen Grundlage der Evolution morphologischer Neuheiten in Solanaceen, mit dem Ziel, zum Verständnis der Subfunktionalisierung bzw. der Neofunktionalisierung der beteiligten Gene beizutragen.

Frühere Studien verließen sich hinsichtlich ihrer Methodik größtenteils auf DNA-Sequenz-Analysen, entbehrten aber der funktionellen Untersuchung duplizierter Gene, was oft zu widersprüchlichen Ergebnissen führte. In dieser Arbeit wurden bioinformatische und experimentelle Herangehensweisen vereint, um die Fragen von Duplikation, Sub- und bzw. oder Neofunktionalisierung, Selektion und Diversifizierung von Genen zu untersuchen, die der - in den Withaninae vorherrschenden - evolutionären Neuerung des Inflated Calyx Syndroms (ICS) zu Grunde liegen.

In der Gattung Withania, ebenso wie in Physalis, nehmen die Sepalen nach der Bestäubung das Wachstum wieder auf und hüllen schließlich die reife Frucht vollständig in einer Ballon-ähnlichen Struktur ein, was als Inflated-Calyx-Syndrom (ICS) bezeichnet wird. Während Withania-Arten eine Vielfalt an aufgeblähten Calyces aufweisen, besitzt Tubocapsicum - eine nahe verwandte Gattung - lediglich einen rudimentären Kelch. Pflanzenhormone - wie Cytokinine und Gibberelline - können die Befruchtungssignale imitieren und sind unerlässlich für die ICS-Ausprägung in Withania. In der diploiden Art *Physalis floridana* ist ein einzelnes Gen – das MADS-Box-Gen *MPF2* – zusätzlich zur ICS-Ausprägung für Blattentwicklung und männliche Fertilität verantwortlich. Dagegen bietet sich in Withania, einer laut Ploidie-Untersuchung offenbar tetraploiden Gattung, die Möglichkeit, diese Funktionen von einander getrennt zu untersuchen. Zwei Klassen von *MPF2*-Orthologen – *MPF2-like-A* (z.B. *WSA206*) und *MPF2-like-B* (z.B. *WSB206*) – wurden aus Withania isoliert, jedoch nur eine Klasse – *MPF2-like-B* (*TAB201*) – aus Tubocapsicum, da *MPF2-like-A*-Gene im Genom dieser Art nicht vorhanden sind. Trotz ähnlicher Gen-Struktur besitzen MPF2-like-A Proteine mit einer 3 Aminosäuren umfassenden Deletion in ihrer C-terminalen Domäne und dem Vorhandensein von 8 zusätzlichen Aminosäuren am C-terminalen Ende im Vergleich zu MPF2 von Physalis einen ungewöhnlichen Aufbau. Nichtsdestotrotz scheinen sie phylogenetisch näher mit MPF2 verwandt zu sein als die MPF2-like-B-Gruppe. Positive Darwinsche Selektion wurde in den phylogenetischen Ästen, die zu Physalis MPF2-like und Withania MPF2-like-A-Proteinen führen, detektiert. Im Gegensatz dazu waren sowohl MPF2-like-Proteine von Arten, die kein ICS aufweisen, als auch alle MPF2-like-B-Proteine unter negativer (reinigender) Selektion.

Während MPF2-like-B-Protein-Überexpression in Arabidopsis-Pflanzen keine phänotypischen Abweichungen im Vergleich zum Wildtyp hervorrief, traten bei MPF2-like-A-Protein-Überexpression stark vergrößerte Sepalen und Abweichungen in Form und Anzahl der Früchte auf, was darauf hindeutet, dass die Funktion beider Proteine unterschiedlich ist.
Dies wird durch ihre divergenten Expressionsmuster gestützt. Während *MPF2-like-A*-Gene in Withania zusätzlich zu vegetativen auch in floralen Geweben exprimiert werden und somit eine Korrelation zur ICS-Bildung aufweisen, ist die Expression von *MPF-2like-B*-Genen auf vegetative Gewebe beschränkt. Dies verhält sich anders in Tubocapsicum, wo *MPF2-like-B*-Gen-Expression in allen untersuchten Geweben gefunden werden konnte, jedoch kein ICS ausgeprägt wird.
Bioinformatische Analysen der Sequenzen von Promotoren und ersten Introns dieser Gene zeigten das Vorhandensein einer konservierten CArG-Box - putatives Bindemotif für AG, SEP1, SEP4, AGL1, and AGL15 - nahe dem Transkriptionsstart in allen hier untersuchten *MPF2-like*-Genen, außer in *MPF2-like-B*-Genen in Withania. Dahingegen treten Auxin-Response-Elemente (AuxREs), die in einer 100 Basenpaare umfassenden hoch-konservierten Region im 1. Intron von *MPF2-like-B*-Genen identifiziert wurden, nicht in *MPF2-like-A*-Genen auf. Sowohl *MPF2-like-A*- als auch *MPF2-like-B*-Gen-Promotoren riefen starke GUS-Expression in den Blüten transgener Arabidopsis-Pflanzen hervor. Dagegen unterdrückten Sequenzen des ersten Introns generell die Reportergen-Expression (GUS und YFP) in Blättern von Arabidopsis bzw. in transient transformierten Blättern von *Nicotiana benthamiana*. Nur MPF2-like-A Proteine sind in der Lage, im Yeast-2-Hybrid-System Homodimere zu bilden. Das Repressor-Motiv am C-Terminus der MPF2-like-A-Proteine ist in Autoaktivierung involviert. Durchmustern einer Arabidopsis-oligo-dT-Hefe-Library mit diesen Proteinen führte zur Identifizierung zahlreicher MADS- und Nicht-MADS-Domänen-Proteine als putative Interaktionspartner. Die Interaktion von MPF2-like-A-Proteinen mit Arabidopsis-Proteinen, die an der Regulation der Blütenentwicklung beteiligt sind (z.B. SOC1, AGL27, AGL19, AGL31, AGL42, AGL68, MAF4, SEP3, SEP4), und die ausschließliche Interaktion von Wihania MPF2-like-B-Proteinen mit Nicht-MADS-Domänen-Proteinen, (z.B. IAA16, EMB1417, ATSIP, AKIN, MYB4), unterstützt eine funktionelle Divergenz der Proteine und korreliert mit ihren divergenten Expressionsmustern. Tubocapsisum MPF2-like-B-Proteine dagegen sind in der Lage mit SOC1, AGL27 und MAF5 zu interagieren, was auf eine intermediäre Funktionalität zwischen den beiden Gruppen von Withania MPF2-like-Proteinen hinweist. Withania MPF2-like-B-Proteine können mit MAGO NASHI-Homologen von Withania, Tubocapsicum und Physalis

interagieren. MAGO NASHI ist ein hoch-konserviertes Protein, das an diversen Entwicklungsprozessen, zum Beispiel Keimzell-Differenzierung, beteiligt ist. Anders als Withania MPF2-like-B-Proteine können MPF2-like-A-Proteine mit keinem dieser MAGO NASHI-Proteine von Solanaceen (WSMAGO1, WSMAGO2, TAMAGO1, TAMAGO2, PFMAGO1, PFMAGO2) interagieren, doch die Fähigkeit von Tubocapsicum MPF2-like-B Proteinen, einige dieser MAGO NASHI-Homologe zu binden, weist erneut auf ihren intermediären Charakter zwischen Withania-MPF2-like-Proteinen hin.

Eindeutig sind *MPF2-like-A*-Gene in die ICS-Ausprägung involviert, die Funktion von *MPF2-like-B*-Genen – abgesehen von einer möglichen Rolle in der Blattentwicklung - bleibt jedoch ungewiss. Allotetraploidisierung trug wesentlich zur Subfunktionalisierung der *MPF2-like*-Gene und positive Darwinsche Selektion ist von großer Bedeutung für die evolutionären Geschichte der Physaleae. Die Ergebnisse dieser Arbeit zeigen die Auswirkungen von Genduplikationen auf die Anpassung und Anpassungsfähigkeit von Organismen und demonstrieren die Leistungsfähigkeit von Studien in evolutionärer Entwicklungsgenetik, die Sequenz-Analyse und funktionellen Untersuchungen kombinieren.

Die hier beschriebenen Ergebnisse sind nicht nur essentielle Grundlage für das Verstehen, wie die evolutionäre Calyx-Neuerung in den Withaninae entstand und erhalten wurde, sondern geben auch ein elegantes Beispiel dafür, wie wilde “evolutionäre Mutanten” zur Aufschlüsselung komplexer Merkmale genutzt werden können.

7 References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W. Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF. 2000a. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J* **24**: 457-466.

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, De Pouplana LR, Martinez-Castilla L, Yanofsky MF. 2000b. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA* **97**: 5328-5333.

Angenent GC, Busscher M, Franken J, Mol JNM, van Tunen AJ. 1992. Differential expression of two MADS-box genes in wild-type and mutant Petunia flowers. *Plant Cell* **4**: 983-993.

Baquar SR. 1967. Cytomorphological studies in the family Solanaceae from West Pakistan. *Genetica* **38**: 388-397.

Becker A, Theissen G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution* **29**: 464-489.

Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J. 2006. Positive selection driving diversification in plant secondary metabolism. *Proc Natl Acad Sci USA* **103**: 9118-9123.

Bielawski JP, Yang Z. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* **59**: 121-132.

Bir SS, Neelam. 1980. Chromosome number reports. *Taxon* **29**: 703-730.

Bonhomme F, Kurz B, Melzer S, Bernier G, Jacmard A. 2000. Cytokinin and gibberellin activate SaMADS A, a gene apparently involved in regulation of the floral transition in *Sinapis alba*. *Plant J* **24**: 103-111.

Borner R, Kampmann G, Chandler J, Gliessner G, Wisman E, Apel K, Melzwe S. 2000. A MADS domain gene involved in the transition to flowering in Arabidopsis. *Plant J.* **24**: 591-599.

Boswell RE, Prout ME, Steichen JC. 1991. Mutations in a newly identified *Drosophila melanogaster* gene, mago nashi, disrupt germ cell formation and result in the formation of mirror-image symmetrical double abdomen embryos. *Development* **113:** 373-384.

Bradley D, Carpenter R, Sommer H, Hartley N, and Coen E. 1993. Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of Antirrhinum. *Cell* **72**: 85-95.

Bryant D, Moulton V. 2004. NeighborNet: an agglomerative algorithm for the construction of planar phylogenetic networks. *Mol Biol Evol* **21**: 255-265.

Carroll S, Grenier JK, Weatherbee SD. 2001. From DNA to diversity: molecular genetics and the evolution of animal design. *Blackwell Publishing* **pp**: 149-171.

Carroll SB. 2005. Evolution at two levels: on genes and form. *Nature* **433**: 481-487.

Clough SJ, Bent A. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**: 735-743.

Coen ES, Meyerowitz EM. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**: 31-37.

D'Arcy WG. 1972. Solanaceae studies II: typification of subdivisions of Solanum. *Ann. Missouri Bot. Gard.* **59**: 262-278.

Danilveskaya ON, Meng X, Selinger DA, Deschamps S, Hermon P, Vansant G, Gupta R, Ananiev EV, Muszynski MG. 2008. Involvement of the MADS-box gene *ZMM4* in floral induction and inflorescence development in *Zea mays*. *Plant Physiology Preview* **10**.1104/pp.107.115261.

D'Arcy WG, Keating RC, Zhang ZY, Peng CI. 2001. The genus Tubocapsicum (Solanaceae). *Botanical Bulletin of Academia Sinica* **42**: 67-84.

D'Arcy WG. 1991. The Solanaceae since 1976, with a review of its biogeography. In: Solanaceae III: taxonomy, chemistry, evolution.—Hawkes JG, Lester RN, Nee M, Estrada N, eds Kew (UK): Royal Botanic Gardens **pp**: 75-137.

Davies B, Egea-Cortines M, de Andrade Silva E, Saedler H, Sommer H. 1996. Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J* **5**: 4330-4343.

Davies B, Motte P, Keck E, Saedler H, Sommer H, Schwarz-Sommer Z. 1999. PLENA and FARINELLI: redundancy and regulatory interactions between two Antirrhinum MADS-box factors controlling flower development. *EMBO J* **18**: 4023-4034.

De Bodt S, Maere S, Van de Peer Y. 2005. Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* **20**: 591-597.

De Bodt S, Theissen G, Van de Peer Y. 2006. Promoter analysis of MADS-box genes in eudicots through phylogenetic footprinting. *Mol Biol Evol* **23**: 1293-303.

de Folter S, and Angenent GC. 2006. *trans* meets *cis* in MADS science. *Trends in Plant Science* **11**: 224-231.

de Folter S, Immink RGH, Kieffer M, (12 co-authors). 2005. Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. *Plant Cell* **17**: 1424-1433.

de Folter S, Shchennikova AV, Franken J, Busscher M, Baskar R, Grossniklaus U, Angenent GC, Immink RG. 2006. A Bsister MADS-box gene involved in ovule and seed development in petunia and Arabidopsis. *The Plant Journal* **47**: 934-946

Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**: 762-765.

Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. 2004. The SEP4 gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* **14**: 1935-1940.

Doebley J, Lukens L. 1998. Transcriptional regulation and the evolution of plant form. *Plant Cell* **10**: 1075-1082.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**: 11-15.

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-15.

Doyle JJ. 1994. Evolution of a plant homeotic multigene family: towards connecting molecular systematics and molecular developmental genetics. *Syst. Biol.* **43**: 307-328.

Duttweiler HM. 1996. A highly sensitive and non-lethal ß-galactosidase plate assay for yeast. *Trends Genet* **12**: 340-341.

Edgar, R, Robert C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 792-797.

Egea-Cortines M, Saedler H, Sommer H. 1999. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *EMBO J* **18**: 5370-5379.

Fan HY, Hu Y, Tudor M, Ma H. 1997. Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant J* **12**: 999-1010.

Force A, Lynch M, Pickett FB, Amores A, YAN YL et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531-1545.

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD. 2000. a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289:** 85-88.

Garcia-Maroto F, Ortega N, Lozano R, Carmona MJ. 2000. Characterization of the potato MADS-box gene STMADS16 and expression analysis in tobacco transgenic plants. *Plant Mol Biol* **42**: 499-513.

Giovannoni J. 2002. A MADS-box gene necessary for fruit ripening at the tomato Ripining-Inhibitor (Rin) locus. *Science* **296**: 343-346.

Goldman N. 1993. Statistical tests of models of DNA substitution. *J Mol Evol* **36**: 182-198.

Gregis V, Sessa A, Colombo L, Kater MM. 2006. AGL24, SHORT VEGETATIVE PHASE, and APETALA1 Redundantly Control AGAMOUS during Early Stages of Flower Development in Arabidopsis. *Plant Cell* **18**: 1373-1382.

Gu Q, Ferrándiz C, Yanofsky MF, Martienssen R. 1998. The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. *Development* **125**: 1509-1517.

Guilfoyle TJ, Hagen G. 2007. Auxin response factors. *Curr Opin Plant Biol* **10**: in press.

Hachet O, Ephrussi A. 2001. Drosophila Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Curr Biol* **11**: 1666-1674.

Hachet O, Ephrussi A. 2004. Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* **428**: 959-963.

Harrison J, Möller M, Langdale J, Cronk Q, Hudson A. 2005. The role of *KNOX* genes in the evolution of morphological novelty in Streptocarpus. *Plant Cell* **17**: 430-443.

Hartmann U, Höhmann S, Nettesheim K, Wisman E, Saedler H Huijser P. 2000. Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. *Plant J.* **21**: 351-360.

Hayes TE, Sengupta P, Cochran BH. 1988. The human c-FOS serum response factor and the yeast factors GRM/pRTF have related DNA-binding specificities. *Genes Dev* **2**: 1713-1722.

He C, Münster T, Saedler H. 2004. On the origin of floral morphological novelties. *FEBS Lett* **567**: 147-151.

He C, Saedler H. 2005. Heterotopic expression of MPF2 is the key to the evolution of the Chinese lantern of Physalis, a morphological novelty in Solanaceae. *Proc Natl Acad Sci USA* **102**: 5779-5784.

He C, Saedler H. 2007. Hormonal control of the inflated calyx syndrome, a morphological novelty, in Physalis. *Plant J* **49**: 935-946.

He C, Sommer H, Grosardt B, Huijser P, Saedler H. 2007. PFMAGO, a MAGO NASHI-Like factor, Interacts with the MADS-domain protein MPF2 from *Physalis floridana*. *Mol Biol Evol* **24**: 1229-241.

He C-Y, Zhang J-S, Chen S-Y. 2002. A soybean gene encoding a proline-rich protein is regulated by salicylic acid, an endogenous circadian rhythm and by various stresses. *Theor Appl Genet* **104**: 1125-1131.

Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theissen G. 2002. Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. *Mol. Biol. Evol.* **19**: 801-814.

Hepper FN. 1991. Old World Withania (Solanaceae): a taxonomic review and key to the species. In J. G. Hawkes, R. N. Lester, M. Nee, and E. Estrada [eds.] Solanaceae III: taxonomy, chemistry, evolution. Royal Botanic Gardens, Kew, UK **pp**: 211-227.

Hepworth SR, Valverde F, Ravenscroft D, Mouradov A, Coupland G. 2002. Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs, *EMBO J* **21**: 4327-4337.

Herrera CM. 2005. Post-floral perianth functionality: contribution of persistent sepals to seed development in *Helleborus foetidus* (Ranunculaceae). *Am. J. Bot.* **92**: 1486-1491.

Hileman LC, Kramer EM, Baum DA. 2003. Differential regulation of symmetry genes and the evolution of floral morphologies. *Proc Natl Acad Sci USA* **100**: 12814-12819.

Honma T, Goto K. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**: 525-529.

Hu J, Saedler H. 2007. Evolution of the inflated calyx syndrome in Solanaceae. *Mol Biol Evol* **24**: 2443-2453.

Huang H, Tudor M, Weiss CA, Hu Y, Ma H. 1995. The Arabidopsis MADS-box gene AGL3 is widely expressed and encodes a sequence-specific DNA-binding protein. *Plant Mol. Biol.* **28**: 549-567.

Huijser P, Klein J, Lönnig WE, Meijer H, Saedler H, Sommer H. 1992. Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene SQUAMOSA in *Antirrhinum majus*. *EMBO J.* **11**: 1239-1249.

Huson DH, and Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254-267.

Huson, DH, Bryant D. 2000. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254-267.

Iqbal M, Datta AK. 2007. Cytogenetic studies in *Withania somnifera* (L.) Dun. (Solanaceae). *CYTOLOGIA* **72**: 43-47.

Jefferson RA, Klass M, Wolf N, Hirsh DJ. 1987. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**: 3901-3907.

Johnson MA, von Besser K, Zhou Q, Smith E, Aux G, Patton D, Levin JZ, Preuss D. 2004. Arabidopsis hapless mutations define essential gametophytic functions. *Genetics* **68**: 971-982.

Kanno A, Saeki H, Kameya T, Saedler H, Theissen G. 2003. Heterotropic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). *Plant Mol. Biol.* **52**: 831-841.

Karol KG, McCourt RM, Cimino MT, Delwiche CF. 2001. The closest living relatives of land plants. *Science* **294**: 2351-2353.

Kataoka N, Yong J, Kim VN, Velazquez F, Perkinson RA, Wang F, Dreyfuss G. 2000. Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* **6**: 673-682.

Kater MM, Colombo L, Franken J, Busscher M, Masiero S, Van Lookeren Campagne MM, Angenent GC. 1998. Multiple AGAMOUS homologs from cucumber and petunia differ in their ability to induce reproductive organ fate. *Plant Cell* **10**: 171-182.

Kaufmann K, Melzer R, Theissen G. 2005. MIKC-type MADS domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* **347**: 183-198.

Kellogg EA. 2004. Evolution of developmental traits. *Curr Opin Plant Biol* **7**: 92-98.

Keys DN, Lewis DL, Selegue JE, Pearson BJ, Goodrich LV, Johnson RL, Gates J, Scott MP, Carroll SB. 1999. Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science* **283**: 532-534.

Kim S, Mizuno K, Fujimura T. 2002. Isolation of MADS-box genes from sweet potatoe (*Ipomoea batatas* (L.) Lam.) expressed specifically in vegetative tissue. *Plant Cell Physiol* **43**: 314-322.

Kim SY, Yu X, Michaels SD. 2008. Regulation of CONSTANS and FLOWERING LOCUS T Expression in Response to Changing Light Quality. *Plant Physiology* **148**: 269-279.

Kim VN, Yong J, Kataoka N, Abel L, Diem MD, Dreyfuss G. 2001. The Y14 protein communicates to the cytoplasm the position of exon-exon junctions. *EMBO J* **20**: 2062-2068.

Knapp S, Bohs L, Nee M, Spooner DM. 2004. Solanaceae – a model for linking genomics with biodiversity. *Comp Funct Genomics* **5**: 285-291.

Knapp S. 2002. Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *J Exp Bot* **53**: 2001-2022.

Koetsier PA, Schorr J, Doerfler W. 1993. A rapid optimized protocol for downward alkaline southern blotting of DNA. *Biotechniques* **15**: 260-262.

Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M. 2003. Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol* **20**: 1963-1977.

Krogan NT, Ashton NW. 2000. Ancestry of plant MADS-box genes revealed by bryophyte (*Physcomitrella patens*) homologues. *New Phytol.* **147**: 505-517.

Lee C, Grasso C, Sharlow MF. 2002. Multiple sequence alignment using partial order graphs. *Bioinformatics* **18**: 452-464.

Lee H, Suh SS, Park E, Cho E, Ahn JH, Kim SG, Lee JS, Kwon YM, Lee I. 2000. The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. *Genes Dev.* **14**: 2366-2376.

Lee H, Suh SS, Park E, Cho E, Ahn JH, Kim SG, Lee JS, Kwon YM, Lee I. 2000. The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. *Genes Dev.* **14**: 2366-2376.

Lee H, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH. 2007. Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes & Dev **21**: 397-402.

Lee PN, Callaerts P, de Couet HG, Martindale MQ. 2003. Cephalopod Hox genes and the origin of morphological novelties. *Nature* **424**: 1061-1065.

Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF. 2000. SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. *Nature* **404**: 766-770.

Liscum E. 2002. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Molecular Biology* **49:** 387-400.

Liu C, Zhou J, Bracha-Drori K, Yalovsky S, Ito T, Yu H. 2007. Specification of Arabidopsis floral meristem identity by repression of flowering time genes. *Development* **134**: 1901-1910.

Llave C, Xie Z, Kasschau KD, Carrington JC. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297**: 2053-2056.

Luo D, Carpenter R, Copsey L, Vincent C, Clark J, Coen E. (I999) Control of organ asymmetry in flowers of Antirrhinum. *Cell* **99**: 367-376.

Luo D, Carpenter R, Vincent C, Copsey L, Coen E. 1996. Origin of floral asymmetry in Antirrhinum. *Nature* **383**: 794-799.

Ma H, Yanofsky MF, Meyerowitz EM. 1991. AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* **5**: 484-495.

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA* **102**: 5454-5459.

Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF. 1992. Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. *Nature* **360**: 273-277.

Masiero S, Li MA, Will I, Hartmann U, Saedler H, Huijser P, Schwarz-Sommer Z, Sommer H. 2004. INCOMPOSITA: a MADS-box gene controlling prophyll development and floral meristem identity in Antirrhinum. *Development* **131**: 5981-5990.

Melzer R, Kaufmann K, Theissen G. 2006. Missing links: DNA-binding and target gene specificity of floral homeotic proteins. *Adv. Bot. Res.* **44**: 209-236.

Merrill ED. 1923. An enumeration of Philippine flowering plants, Vol. 3. Manila, **pp**: 424

Michaels SD, Amasino RM. 1999. Flowering Locus C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**: 949-956.

Micklem DR, Dasgupta R, Elliott H, Gergely F, Davidson C, Brand A, Gonzalez-Reyes A, St Johnston D. 1997. The mago nashi gene is required for the polarisation of the oocyte and the formation of perpendicular axes in Drosophila. *Curr Biol* **7**: 468-478.

Miller W. 2005. Mulan: Multiple sequence local alignment and visualization for studying function and evolution. *Genome research* **15**: 184-194.

Mohr SE, Dillon ST, Boswell RE. 2001. The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize oskar mRNA during Drosophila oogenesis. *Genes Dev* **15**: 2886-2899.

Molnar A, Csorba T, Lakatos L, Varallyay E, Lacomme C, Burgyan J. 2005. Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. *J. Virol.* **79**: 7812-7818.

Mondragón-Palomino M. Theissen G. 2008. MADS about the evolution of orchid flowers. *Trends Plant Sci* **13**: 51-59.

Münster T, Deleu W, Wingen LU, Ouzunova M, Cacharrón J, Faigl W, Werth S, Kim JTT, Saedler H, Theissen G. 2002a. Maize MADS-box genes galore. *Maydica* **47**: 287-301.

Münster T, Faigl W, Saedler H, Theissen G. 2002b. Evolutionary aspects of MADS-box genes in the eusporangiate fern Ophioglossum. *Plant Biol.* **4**: 474-483.

Münster T, Pahnke J, Rosa AD, Kim JT, Martin W, Saedler H, Theissen G. 1997. Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc. Natl. Acad. Sci. USA* **94**: 2415-2420.

Murashige T, Skoog F. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol Plant* **15**: 473-497.

Nagasawa N. et al. 2003. SUPERWOMAN1 and DROOPING LEAF genes control floral organ identity in rice. *Development* **130**: 705-718.

Negi MS, Singh A, Lakshmikumaran M. 2000. Genetic variation and relationship among and within Withania species as revealed by AFLP markers. *Genome* **43**: 975-980.

Newmark PA, Boswell RE. 1994. The mago nashi locus encodes an essential product required for germ plasm assembly in Drosophila. *Development* **120**: 1303-1313.

Newmark PA, Mohr SE, Gong L, Boswell RE. 1997. mago nashi mediates the posterior follicle cell-to-oocyte signal to organize axis formation in Drosophila. *Development* **124**: 3197-3207.

Ng M, Yanofsky F. 2001. Function and evolution of the plant MADS-box gene family. *Nat Rev Gen* **2**: 186-195.

Nylander JA. 2004. MrModelTest, 2.0th edition. Upsalla, Distributed by the author. Evolutionary Biology Centre, Uppsala University.

Ohno S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin, Germany.

Olmstead RG, Bohs L. 2007. A summary of molecular systematic research in Solanaceae: 1982-2006. In: D.M. Spooner et al. (Hrsg.): Solanaceae VI: Genomics Meets Biodiversity. *Acta Horticulturae* **745**: 255-268.

Olmstead RG, Kim KJ, Jansen RK, Wagstaff SJ. 2000. The phylogeny of the Asteridae sensu lato based on chloroplast ndhF gene sequences. *Mol Phylogenet Evol* **16**: 96-12.

Olmstead RG, Palmer JD. 1997. Implications for the phylogeny, classification, and biogeography of Solanum from cpDNA restriction site variation. *Syst Bot* **22**: 19-29.

Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer JD. 1999. Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. In: Solanaceae IV: advances in biology and utilization—Nee M, Symon DE, Lester RN, Jessop JP, eds. Kew (UK): Royal Botanic Gardens **pp**: 111-137.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet* **34**: 401-437.

Ovcharenko I, Loots GG, Giardine BM, Milnmei H, Jian MA, Hardison RC, Stubbs L,

Ozga JA, Reinecke DM. 2003. Hormonal interactions in fruit development. *J Plant Growth Regul* **22**: 73-81.

Pagnussat GC, Yu HJ, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie LF, Ye D, Sundaresan V. 2005. Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development* **132**: 603-614.

Parenicova L, de Folter, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L. 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis; New openings to the MADS world. *Plant Cell* **15**: 1538-1551.

Paz-Ares J, Ghosal D, Saedler H. 1990. Molecular analysis of the C1-I allele from *Zea mays*: a dominant mutant of the regulatory C1 locus. *EMBO J.* **9**: 315-321.

Pogson GH, Mesa KA. 2004. Positive Darwinian selection at the pantophysin (*Pan I*) locus in Marine Gadid fishes. *Mol Biol Evol* **21**: 65-75.

Prince VE, Pickett FB. 2002. Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet* **3**: 827-837.

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF. 1995. Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* **140**: 345-356.

Rensing SA, Lang D, Zimmer AD. **et al. 2007.** The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**: 64-69.

Riechmann JL, Wang MQ, Meyerowitz EM. 1996. DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucl Acids Res* **24**: 3134-3141.

Riechmann, JL, Meyerowitz EM. 1997. MADS domain proteins in plant development. *Biological Chemistry* **378**: 1079-1101.

Riese M, Faigl W, Quodt V, Verelst W, Matthes A, Saedler H, Münster T. 2004. Isolation and characterization of new MIKC*-type MADS-box genes from the moss *Physcomitrella patens*. *Plant Biol* **7**: 307-314.

Rijpkema AS, Gerats T, Vandenbussche M. 2007. Evolutionary complexity of MADS complexes. *Current Opinion in Plant Biology* **10**: 32-38.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.

Rose AB. 2004. The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis. *Plant J* **4**: 744-751.

Rose AB. 2008. Intron-mediated regulation of gene expression. *Curr Top Microbiol Immunol* **326**: 277-290.

Rounsley SD, Ditta GS, Yanofsky MF. 1995. Diverse roles for MADS box genes in Arabidopsis development. *Plant Cell* **7**: 1259-1269.

Sambrook J, Russell DW. 2001. Molecular cloning: a laboratory manual. Third edition. Cold Spring Harbor Laboratory Press**.**

Schwarz-Sommer Z, Gierl A, Cuypers H, Peterson PA, Saedler H. 1985. Plant transposable elements generate the DNA sequence diversity needed in evolution. *EMBO J.* **4**: 591-597.

Schwarz-Sommer Z, Huijser P, Wolfgang N, Saedler H, Sommer H. 1990. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **250**: 931-936.

Schwarz-Sommer Z. 1992. GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of Antirrhinum floral organogenesis. *EMBO J.* **11**: 4693- 4704.

Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards WJ, Peacock WJ, Dennis ES. 1999. The FLF MADS-box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. *Plant Cell* **11**: 444-458.

Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES. 2000. The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC*). Proc. Natl. Acad. Sci. USA* **97**: 3753-3758.

Shore P, Sharrocks AD. 1995. The MADS-box family of transcription factors. *Eur J Biochem* **229**: 1-13.

Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **99**: 13627-13632.

Singer T, Gierl A, Peterson PA. 1998. Three new dominant *C1* suppressor allels in *Zea mays*. *Genet. Res. Camb.* **71**: 127-132.

Sommer H, Beltran JP, Huijser P, Pape H, Lonni, W-E, Saedler H, Schwarz-Sommer Z. 1990. Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J* **9**: 605-613.

Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo JM, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U. 2007. Positive Darwinian selection at the imprinted *MEDEA* locus in plants. *Nature* **448**: 349-352.

Swidzinski JA, Zaplachinski ST, Chuong SD, Wong JF, Muench DG. 2001. Molecular characterization and expression analysis of highly conserved rice mago nashi1 homolog. *Genome* **44**: 394-400.

Swofford DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland MA. (Program).

Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, Münster T, Theissen G, Nozaki H, Ito M. 2005. Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. *Proc Natl Acad Sci USA* **102**: 2436-2441.

Tandre K, Albert VA, Sundas A, Engström P. 1995. Conifer homologues to genes that control floral development in angiosperms. *Plant Mol. Biol.* **27**: 69-78.

Tautz D. 2000. Evolution of transcriptional regulation. *Curr Opin Genet Dev* **10**: 575-579.

Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter K-U, Saedler H. 2000. A short history of MADS-box genes in plants. *Plant Mol Biol* **42**: 115-149.

Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, Saedler H. 2000. A short history of MADS-box genes in plants. *Plant Mol Biol* **42**: 115-149.

Theissen G, Kim JT, Saedler H. 1996. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J Mol Evol* **43**: 484-516.

Theissen G, Saedler H. 1995. MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. *Curr Opin Genet Dev* **5**: 628-639.

Theissen G, Saedler H. 1999. The golden decade of molecular floral development (1990–1999): a cheerful obituary *Dev. Genet.* **25**: 181-193.

Theissen G, Saedler H. 2001. Plant biology: floral quartets. *Nature* **409**: 469-471.

Theissen G. 2001. Development of floral organ identity: stories from the MADS house. *Curr Opin Plant Biol* **4**: 75-85.

Theissen G. 2006. The proper place of hopeful monsters in evolutionary biology. *Theory Biosci* **124**: 349-369.

Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönnig WE, Saedler H, Sommer H, True JR, Carroll SB. 2002. Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol*. **18**: 53-80.

Tsuchimoto S, van der Krol AR, Chua NH. 1993. Ectopic expression of pMADS3 in transgenic petunia phenocopies the petunia blind mutant. *Plant Cell* **5**: 843-85.

Ulmasov T, Hagen G, Guilfoyle TJ. 1999a. Activation and repression of transcription by auxin-response factors. *Proc Natl Acad Sci USA* **96**: 5844-5849.

Ulmasov T,Hagen G, Guilfoyle TJ. 1999b. Dimerization and DNA binding of auxin response factors. *Plant J* **19**: 309-319.

Valido AA, Manuel NB, Medina FM. 2003. Fleshy fruits in the diet of Canarian lizards *Gallotia galloti* (Lacertidae) in a xeric habitat of the island of Tenerife. *Journal of Herpetology* **pp**: 741-747.

van Tunen AJ, Eikelboom W, Angenent GC. 1993. Floral organogenesis in Tulipa. *Flow Newl* **16**: 33-38.

Vandenbussche M, Theissen G, Van de Peer Y, Gerats T. 2003a. Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Research* **31**: 4401-4409.

Vandenbussche M. et al. 2003b. Toward the analysis of the petunia MADS box gene family by reverse and forward transposon mutagenesis approaches: B, C and D floral organ identity functions require SEPALLATA-like MADS box genes in petunia. *Plant Cell* **15**: 2686-2693.

Veley KM, Michaels SD. 2008. Functional redundancy and new roles for genes of the autonomous floral-promotion pathway. *Plant Physiology* in press.

Verelst W, Saedler H, Munster T. 2007. MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters. *Plant Physiology* **143**: 447-460.

Vergara-Silva F. 2003. Plants and the Conceptual Articulation of Evolutionary Developmental Biology. *Biology and Philosophy* **18**: 249-284.

Vivian F, Irish I. 2005. Gene duplication, diversification and redeployment. *Flower development and evolution* **15**: 454-460.

Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, Schuch W, Giovannoni J. 2002. A MADS-box gene necessary for fruit ripening at the tomato Ripining-Inhibitor (Rin) locus. *Science* **296**: 343–346.

Webster MA, Gilmartin PM. 2003. A comparison of early floral ontogeny in wild-type and floral homeotic mutant phenotypes of Primula. *Planta* **216**: 903-917.

Weigel D, Meyerowitz EM. 1994. The ABCs of floral homeotic genes. *Cell* **78**: 203-209.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Mol Biol* **42**: 225-249.

West et al. 1998. DNA binding and dimerisation determinants of Antirrhinum majus MADS-box transcription factors. *Nucleic Acids Res*. **26**: 5277-5287.

Weyers JDB, Paterson NW. 2001. Plant hormone and the control of physiological processes. *New Phytol* **152**: 375-407.

Winter KU, Becker A, Münster T, Kim JT, Saedler H, Theissen G. 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. USA* **96**: 7342-7347.

Yamaguchia T, LeebGY, Miyaoc A, Hirochikac H, Anb G, Hiranoa HY. 2006. Functional Diversification of the Two C-Class MADS Box Genes OSMADS3 and OSMADS58 in *Oryza sativa*. *The Plant Cell* **18**: 15-28.

Yang Z, Bielawski B. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution* **5**: 496-503.

Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol* **46**: 409-418.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* **17**: 32-43.

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* **19**: 908-917.

Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **22**: 1107-1118.

Yanofsky MF, Ma H, Bowman MH, Drews JN, Feldmann KA, Meyerowitz EM. 1990. The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* **346**: 35-39.

Yu H, Ito T, Wellmer F, Meyerowitz EM. 2004. Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development. *Nat Genet* **36**: 157-161.

Yu Y, Xu EL, Tan, Kumar PP. 2002b. AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. *Proc. Natl. Acad. Sci. USA* **99**: 16336–16341.

Zhang XM, Dao ZL, Long CL, Li H. 2006. Karyotypical studies on an endangered species *Alcimandra cathcartii* (Magnoliaceae). *Acta Botanica Yunnaica* **28**: 268-270.

Zhang Z, Lu A, D'Arcy WG. 1994. Solanaceae. In: Flora of China, Vol. 17 (Verbenaceae through Solanaceae)—Wu Z, Peter HR, eds. Beijing: Science Press; St Louis: Missouri Botanical Garden Press. **pp**: 300-332.

Zhao XF, Nowak NJ, Shows TB, Aplan PD. 2000. MAGO NASHI interacts with a novel RNA-binding protein. *Genomics* **63**: 145-148.

8 Supplement

8.1 Supplementary figures

Figure S1: *MPF2-like-A* genes have differentiated into *MPF2-like-A*- and *–B* genes

ClustawlW alignment of the full-length cDNA sequences of *MPF2-like-A* and *MPF2-like-B* genes of *Withania somnifera* are depicted. The 5' and 3'-UTRs are shown in blue colour. Black and blue solid boxes indicate the start and stop codons for both the genes, respectively. MADS-box is underlined with black stars; I-box is underlined with green stars; K-box is underlined with blue stars; C-terminal region is underlined with red stars. Red solid boxes highlight the 9 base pair deletion and 24 base pair extension in the C-terminal regions of *MPF2-like-A* gene, respectively. ClustalW alignment parameters are given as a text.

```
ClustalW (v1.4) multiple sequence alignment

2 Sequences Aligned           Alignment Score = 6422
Gaps Inserted = 18            Conserved Identities = 909

Pairwise Alignment Mode: Slow
Pairwise Alignment Parameters:
    Open Gap Penalty = 10.0   Extend Gap Penalty = 0.1

Multiple Alignment Parameters:
    Open Gap Penalty = 10.0   Extend Gap Penalty = 0.1
    Delay Divergent = 40%     Transitions: Weighted

Processing time: 1.4 seconds


WSA206      1                                                      0
WSB206      1 AAAACGATAGTAGCTCTTTCAATTCTCTCTCTCTCTAGGTTTTAATTGGT   50


WSA206      1               AATATCTTCCAAC-----AACAAATTAAAAATTCTT   31
WSB206     51 GGAGCTCTCCAATAATTATCTTCCTACTATACAACAAATTAAAAAATCTT 100
                           * ******** **     ************* ****

WSA206     32 ATTCTCTCTCTTTCCTTCTTCAATACCAAGAACAACAAG-TGAACAGTGC   80
WSB206    101 ----TCT-----TCTTTCTTCAACTCCAAGAACAAGAAACTGAACCGTTC  141
                  ***     ** ********  ********** **  ***** ** *

WSA206     81 TGATTCCA--TCTCCAAATACTATCAAAATTAGGAAAAATTGAAGGGGGA  128
WSB206    142 TCCTATAAGGTTTGTTGAAGCAGCTAAGCGGGGGAACAATTAGTAATATC  191
               *  *   *  * *    *  *    **     **** ****

WSA206    129 TCAT-TAAT-----ATGGCAAGAGAGAAGATTAAGATAAAGAAGATAGAT  172
WSB206    192 TTGTGTAATTCAGAATGGCAAGGGAGAAGATCAAGATACAGAAGATAGAT  241
               *  * ****     ******** ******** ****** **********
```

```
WSA206   173 AATATAACAGCAAGGCAAGTGACATTTTCAAAGAGGAGAAGAGGGATTTT  222
WSB206   242 AATATAACAGCAAGACAAGTGACATTTTCAAAGAGGAGAAGAGGGCTTTT  291
             ************** ****************************** ****

WSA206   223 CAAGAAAGCCGAAGAGCTTTCTGTTCTTTGTGATGCTGATGTTGCTCTTA  272
WSB206   292 CAAGAAAGCTGAAGAGCTTTCTGTTCTTTGTGATGCTGATGTTGCTATTA  341
             ********* ********************************** ***

WSA206   273 TTGTTTTCTCTGCTACTGGCAAACTCTTTGAATTTGCTAGCTCCAGCATG  322
WSB206   342 TTATTTTCTCTGCTACTGGCAAACTCTTTGAATATGCTAGCTCCAGCATG  391
             ** ****************************** ****************

WSA206   323 AAGGATATCCTTGGAAAATATAAGTTGCAGTCAGCTAACCTTGAGAAAGT  372
WSB206   392 AGGGATATCCTTGGAAAATATAAGTTGCAGTCAGCTAATCTTGAGAAAGT  441
             * ********************************** ***********

WSA206   373 TTACCAACCTTCCCTTGACTTACAGCTAGAGAATAGCCTCAACGTGAGAT  422
WSB206   442 TGACCAGCCTTCCCTTGACTTACAGCTAGAGAATAGCCTCAACATGAGAT  491
             * **** **************************************** *******

WSA206   423 TATGCAAGCAGGTAGCTGATAAAACTCGTGAGCTCAGGCAGATGAAAGGT  472
WSB206   492 TAAGCAAGCAGGTAGCTGACAAAACTCGTGAGCTAAGGCAGATGAGAGGT  541
             ** ***************** ************** ********** ****

WSA206   473 GAGGAACTTGAAGGATTGAGCTTAGAAGAATTACAACAACTTGAGAAAAG  522
WSB206   542 GAGGAACTTGAAGGATTGAGTTTAGAAGAATTACAACAAATTGAGAAAAG  591
             ******************** ****************** **********

WSA206   523 ACTTGAAGCTGGATTCAACCGTGTGGTTGAGATTAAGGATAGACGAATTA  572
WSB206   592 ACTTGAAGCTGGATTCAACCGTGTGCTTGAGATTAAGGGTGCACGAATTA  641
             ************************** *********** *  ********

WSA206   573 TGGATGAAATTGCCAACCTTCAAAGAAAGGGTGCTGAGCTGATGGAAGAA  622
WSB206   642 TGGATGAAATTACCAAACTTCAAAGAAAGGGTGCTGAGCTGATGGAAGAA  691
             *********** **** ********************************

WSA206   623 AACAAACAATTGAAAGAGAAAATGGAAATGACGAAGGTAGGGAAGTTGCC  672
WSB206   692 AAAAAACAATTGAACCAGAAAATGGAAATGTTGAAGGAAGGAAAGTTGCC  741
             ** ***********  **************  ***** *** ********

WSA206   673 CTTACTAACAGACATGG---------TGATGGAAGAGGGACAATCATCGG  713
WSB206   742 CTTAGTAATAGACATGGATTGCATGGTGATGGAAGAAGGCCAATCATCTG  791
             **** *** ********         ********** ** ******** *

WSA206   714 ACTCTATAATTACCACCAATAATATCTGCAGTAGTAATAGTGGTCCTCCT  763
WSB206   792 AGTCTATAATTACCACAAATAATGTCTGCAGTAGTAATAGTGGTCCTCCT  841
             * ************** ****** ***************************

WSA206   764 CCTGAAGATGATTGCTCAAATGTTTCTTTAAAGTTAGGTTATAACAATGG  813
WSB206   842 CCTGAAGATGATTGCTCAAATGCTTCTTTGAAGCTAGGTTGTAACAATGG  891
             ********************** ****** *** ****** *********

WSA206   814 TCCTGCAGCAGTGGATGATGATTGCTCAATTACATCTTTGACGTTAGGGC  863
WSB206   892 TCCTGCAGCTGTGGAGGATGATAGCTCAATTACATCTTTGAAATTAGGGT  941
             ********* ***** ****** ******************  ******
```

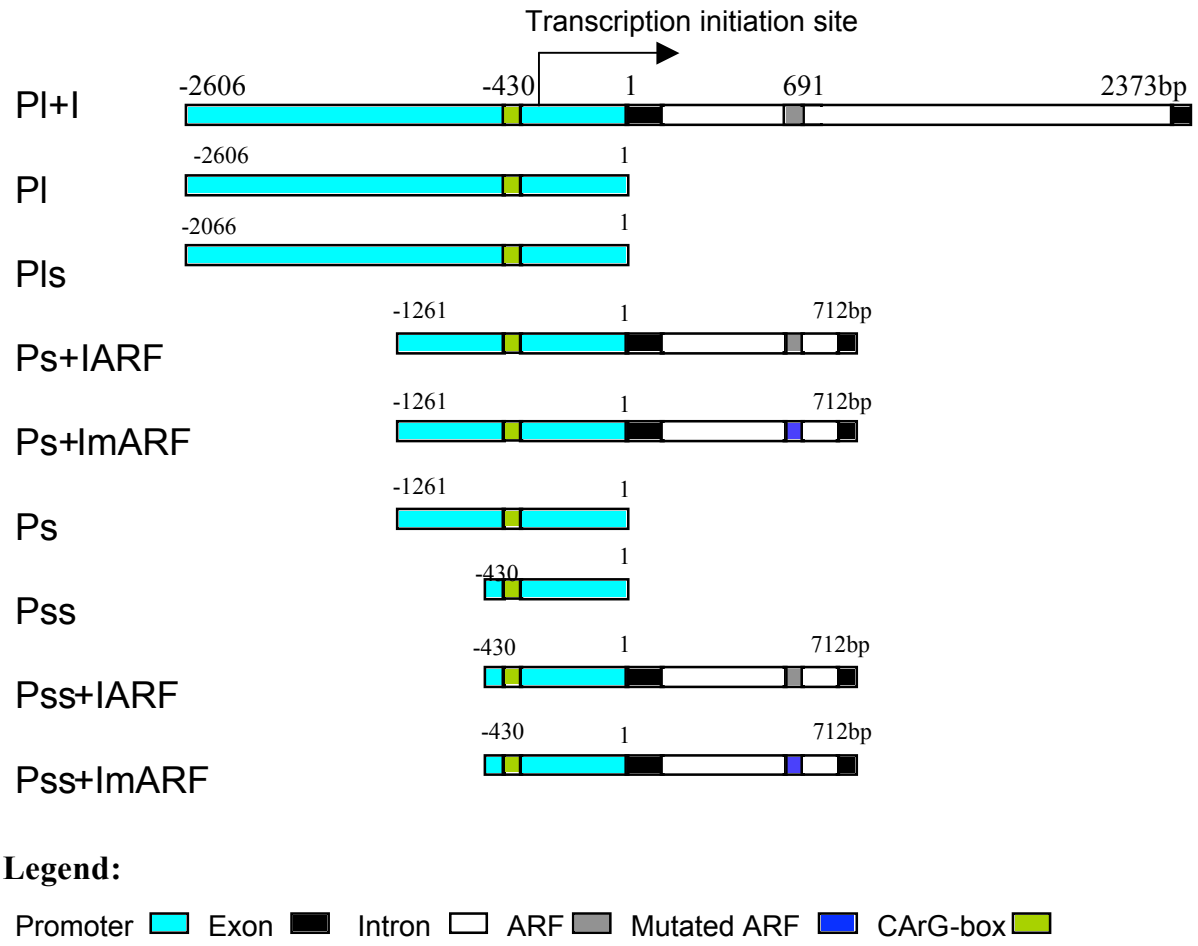
```
WSA206   864 AAGGCTACCACTCAGCTAGCAAGGGAAGATTGTACTAATAACTTTTGGAT  913
WSB206   942 -----TATCATTCAGCTAACTA------ACTCAATGAAGGAAGCTTTTGT  980
                  ** ** ******* * *      * *  *  **  *   **   *

WSA206   914 GGAGACACAACTTACATGTAACTGTCAGTAGAATATAATTGTCATCACTT  963
WSB206   981 TGAAGAATAATTT---------TGGGATGAAGAAATAATTATCAGTGCTT 1021
              **   * ** **         **  *  *  * ****** ***   ***

WSA206   964 GTACTATGTATTATTATGTATGGTGGAGACAACTTGAGAAATGACATAAT 1013
WSB206  1022 GTACTATGTATTAT---GTATGCATGGGACAACCTAAGAAAGGGAAAAAT 1068
             ***************   *****   * ****** * ***** *  * ***

WSA206  1014 CC-------TCATGAGT-TCAATTCGTGTACT-----------AATGTGG 1044
WSB206  1069 TCAAATTAATCAAGAGTCTCATTTTGGGTCTTTGGCAATCTACAGTGTAA 1118
              *       *** **** *** ** * **  *           * ***

WSA206  1045 ---------------------------TGTTTTTTGTATTTTTGATATGT 1067
WSB206  1119 CTTATAAGTCTACATGACTTTGATAGTTATTCTCAAAATCTTGTATATGT 1168
                                       * ** *    ** **  ******

WSA206  1068 ATT------TAATGTGGTGTCACCAA----------TTAGAAAAAAAAAA 1101
WSB206  1169 ATTATAAACTTATGGAATATTATTAAATTGATACAGTTAAAAAAAAAAAA 1218
             ***      * ***   * * *  **          *** **********

WSA206  1102 AAAAAA     1107
WSB206  1219 AAAAAAAAAA 1228
```

Figure S2:

Scheme of making *MPF2-like* promoter:GUS/YFP and promoter-1st exon-1st intron:GUS/YFP constructs

In the scheme only 9 different constructs of variable lengths corresponding to *WSA206* gene are shown. The same scheme applies to *WSB206*, *TAB201* and *MPF2* except two missing constructs in the case of *MPF2* (Pl and Pls). For *V001* five constructs were designed (Pl+I, Pl, and Pls). Numbers indicate the upstream and down stream positions of *cis*-elements as well as the length of promoters and introns in base pair. The legend is also given. For more details see Table 3.5 the Results part.

P = promoter; l = long; I = 1st exon-1st intron; s = short; ss = very short; m = mutated, 1 = first base of ATG.



Legend:

8.2 Supplementary tables

Table S1: Some of the important oligonucleotide primers used in the study

3' RACE *MPF2-like* genes:

K003-F 5'-ATGGCAAGAGAGAAGATTAAGATAAAGAAGA-3'

K004-R 5'-GACCACGCGTATCGATGTCGAC-3'

3' RACE *MPF1-like* genes:

K001-F 5'-ATGGTGAGACAAAAATCCAGATCAAGAAGATAGC-3'

K002-R 5'-GACCACGCGTATCGATGTCGAC-3'

Isolation of genomic loci for *WSA206*:

K042-F 5'-TTCTTATTCTCTCTCTTTCCTTCTTCAATACCA-3'

K043-R 5'-GTACACGAATTGAACTCATGAAGGATTATGT-3'

Isolation of genomic loci for *WSB206*:

K044-F 5'-CGATAGTAGCTCTTTCAATTCTCTCTCTCTCT-3'

K045-R 5'-CTGTAGATTGCCAAAGACCCAAATGAGAC-3'

Isolation of genomic loci for *TAB201*:

K046-F 5'-AGTTCCAAGGACAAGAAAGTGAACAGTGCTGATT-3'

K047-R 5'-GAGAATAATTATCAAAGTCACGTGAACTC-3'

5' RACE and promoter isolation of *MPF2-like* genes:

K022-R 5'-CAGAAAGCTCTTCAGCTTTCTTAAAAAGCC-3'

K023-R 5'-CTCTTCTCCTCTTTGAAATGTCACTTGTC-3'

K024-R 5'-CAGCATCACAGAAACAGATAGCTCTTCAGC-3'

RT-PCR of *MPF2-like-A* (*WSA206*) gene expression:

K031-F 5'-TCAGGCAGATGAAAGGTGAGG-3'

K032-R 5’-GACCTTCTTCCATCACCATGCAATCC-3’

RT-PCR of *MPF2-like-B* (*WSB206* and *TAB201*) gene expression:

K033-F 5’-CGATGATTGTCCCTCTTCCATCACC-3’

K034-R 5-GACCTTCTTCCATCACCATGCAATCC-3’

18S rDNA

Ji221-F 5’-AAAGGAATTGACGGAAGGGCA-3’

Ji222-R 5’-AGCTAGTAGCAGGCTGAGGT-3’

RT-PCR of transgenic Arabidopsis lines over-expressing *MPF2-like* genes:

K003-F 5’-ATGGCAAGAGAGAAGATTAAGATAAAGAAGA-3’

K027-R 5’-CATCTTCAGGAGGAGGACCACTATTAC-3’

Isolation of *atpB* chloroplast marker:

Ji101-F 5’-GTTTACGGTCAGATGAATG-3’

Ji012-R 5’-CATTCATCTGACCGTATTA-3’

Isolation of *matK* chloroplast marker:

Ji013-F 5’-CTATATCCACTTATCTTTCAGGAG-3’

Ji014-R 5’-GGAACAAGGGTATCGAACTTCTTAA-3’

***MPF2-like-A* gene (*WSA206*) probe for Southern hybridization:**

K122-F 5’-GCCCTTAACTAACAGACATGGTGATG-3’

K123-R 5’-GTCTCCATCCAAAAGTTATTAGTACAATC-3’

***MPF2-like-B* gene (*WSB206*) probe for Southern hybridization:**

K124-F 5’-CCTTAGTAATAGACATGGATTGCATGGTG-3’

K125-R 5’-CTTCAACAAAGCTTCCTTCATTGAGTTAG-3’

WSA206 **gene constructs for yeast 2-hybrid and over-expression in Arabidopsis:**

RK64-F 5’-GCTCCATGGCAAGAGAGAAGATTAAGATAAA-3’

K187-R 5’-GCTGGATCCAGTACAATCTTCCCTTGCTAGC-3’

WSB206 **gene constructs for yeast 2-hybrid and over-expression in Arabidopsis:**

RK65-F 5’-GCTCCATGGCAAGGGAGAAGATCAAGATACA-3’

K188-R 5’-GCTGGATCCAGTACAATCTTCCCTTGCTAGC-3’

TAB201 **gene constructs for yeast 2-hybrid and over-expression in Arabidopsis:**

RK66-F 5’-GCTCCATGGCAAGAGAGAAGATCAAGATAAA-3’

K189-R 5’-GCTGGATCCGCTGAATGATAGCCCTAATTTC-3’

Isolation of ***MAGO NASHI-like*** **genes from Withania:**

K263-F 5’-GTCCCATGGGGGAATTGGAAGAGAATGGGAAT-3’

K265-R 5’-GTCGGATCCTTAAATGGGTTTGATCTTGAAATGGAG-3’

Isolation of ***MAGO NASHI-like*** **genes from Tubocapsicum:**

K264-F 5’-GTCCCATGGATGGGGGAGATGGCAGAGAACGAGGA-3’

KK265-R 5’-GTCGGATCCTTAAATGGGTTTGATCTTGAAATGGAG- 3’

Over-expression of ***35S:WSA206:YFP*** **constructs in Arabidopsis:**

K181-F 5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCAAGAG
AAGAATTAAGATAA-3’

K184-R 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGTACAATCTT
CCCTTGCTAGCTGA-3’

Over-expression of ***35S:WSB206:YFP*** **constructs in Arabidopsis:**

K182-F 5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCAAGGG
GATCAAGATACAGAA-3’

K185-R 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGCTGAATGAT

AACCCTAATTTCAA-3’

Over-expression of *35S:TAB201:YFP* constructs in Arabidopsis:

K183-F 5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCAAGAG AGAAAGATCAAGATAA-3’

K186-R 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGCTGAATGAT AGCCCTAATTTCAA-3’

Primers obtained from different kits have not been mentioned in this table. In addition, primers used in making GUS, YFP and yeast constructs and their derivatives have also not been listed as well. Detailed informations on primer sequences and PCR conditions have been deposited in Saedler/Münster group, MPIZ.

8.3 Abbreviations

al.	alii
ARF	auxin response factor
AuxREs	auxin response factor elements
6-BAP	6-benzylaminopurine
BLAST	Basic Local Alignment Search Tool
CArG	C A/T-rich G motif
cDNA	complementary DNA
c-type	classic-type
*-type	star-type
DNA	deoxyribonucleic acid
ω	omega
Δ	delta
Fig.	Figure
FSS	foliose sepal syndrome
GA3	gibberellic acid-3
GUS	ß-Glucuronidase
IAA	indole-3-acetic acid
ICS	inflated calyx syndrome
2-IP	2-isopentenyladenine
kb	kilobase
LSCM	laser scanning confocal microscope
M	molar
MADS	acronym for the four founder proteins MCM1 (*S. cerevisiae*), AGAMOUS (*A. thaliana*), DEFICIENS (*A. majus*) and SRF (*H. sapiens*)
MIKC	domain structure consisting of MADS, intervening, keratin-like and C-terminal domain
ML	maximum likelihood
MP	maximum parsimony
MPF2	MADS-box gene number 2 from *Physalis floridana*
MS	Murashige and Skoog salt mixture
Mulan	multiple sequence local alignment
mYFP	monomeric yellow fluorescent protein

NCBI	national centre for biotechnology information
ORF	open reading frame
PAML	phylogenetic analysis by maximum likelihood
PCR	polymerase chain reaction
5'RACE	5' rapid amplification of cDNA ends
RAGE	rapid amplification of genomic DNA ends
R-motif	repressor motif
rpm	rounds per minute
SD	standard dropout
SRE	serum response element
RT-PCR	reverse transcription PCR
TFBS	transcription factor binding sites
UTR	un-translated region
WT	wild type
Y2-H	yeast two-hybrid
YFP	yellow fluorescent protein

8.4 Acknowledgments

First and foremost I take this opportunity to offer my reverent and sincerest gratitude to my supervisor, Prof. Dr. Heinz Saedler, who not only gave me this interesting topic about functional evolution but has supported me throughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way at Max Planck Institute for Breeding Research, Cologne, Germany.

I ardently extend my gratitude to Prof. Dr. Wolfgang Werr for acting as a co-referee for my thesis evaluation.

I am thankful to Prof. Dr. Ulf-Ingo Flügge for chairing the session and Dr. Gregor Schmitz for writing the protocols during thesis defence.

My gratefulness goes to Dr. Thomas Münster for giving me the opportunity to work in his lab. I attribute the level of my PhD degree to his encouragement and effort and without him this thesis, too, would not have been completed or written. One simply could not wish for a better or friendlier group leader.

Sincere appreciation and gratitude to Prof. Dr. Chaoying He, co-group leader, for his help in lab experiments, valuable comments, skilful suggestions, keen interest and constructive criticism which helped a lot to improve the quality of work.

The completion of this study is due to idea, collaborative work, generous help and guidance by Dr. Jinyong Hu.

In the various laboratories, my colleagues Britta Grosardt and Wolfram Faigl have aided me for many years in running the equipment and learning useful techniques. Wolfram Faigl has offered much advice and insight throughout my work in the lab.

I have been always blessed with friendly assistance of Dr. Elmon Schmelzer in Confocal Laser Scanning Microscopy. His continued patience and unfailing attention has graced this manuscript with fine images. The efforts of Dr. Peter Huijser in normal microscopy and photography are duly acknowledged.

In my daily work I have been blessed with a friendly and cheerful group of fellow students. Thanks to Simone Riss not only for collaborative work, also for useful discussions on a variety of topics including research topic as well. Michiel Kwantes kept me on my theoretical toes with his questions, provided good suggestions and helped in lab work. Daniela Liebsch provided good arguments and constructive suggestions particularly in writing of this manuscript and often made sure none of us starved and went thirsty.

I always found my seniors colleagues, Dr. Wim Verelst and Dr. Oliver Zobell willing for help. I am also grateful to other colleagues in the department and MPIZ for useful discussions and help.

My thankfulness goes to Dr Richard G.H. Immink, Plant Research International Bioscience, Plant Developmental Systems (PDS), Wageningen, the Netherlands, for providing me the yeast constructs and Dr Michael Goodin, Department of Plant Pathology, University of Kentucky, Lexington USA, for giving me RFP and YFP vector constructs.

The ADIS and SUSAN teams are acknowledged for their professional services.

I am grateful to my friends, Dr Ali Ahmad Naz, Muhammad Ilyas, Muhammad Usman, Sajid Hussain, Abid Mehmood Alvi, Muhammad Hasnain Raza and Dr Abdul Wakeel for their help, amicable behaviour and cheerful company throughout my studies.

How can I ignore to express my appreciation and deep sense of gratitude from the citadel of my heart to my sweet deceased parents, who remember me in their prayer and let me spare to complete my studies for a long time, when alive. My ineffable indebtedness is due to my wife Ruqia, my children Ayesha, Aida and little Minnu for allowing me to spare more time in the lab than at home.

Last but not least the Higher Education Commission (HEC) of Pakistan, German Academic Exchange Service (DAAD) and Max Planck Society are acknowledged for providing funds.

Finally, I feel greatly obliged to Christiane Wojtera, whom always I found ready for solving many problems beyond lab work necessary for smooth stay of my family here in Cologne. Germany.

8.5 Eidesstattliche Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Heinz Saedler betreut worden.

8.6 Lebenlauf

Name	Muhammad Ramzan Khan
Adresse	Kolibriweg 14, 50829, Köln
Geburtsdatum	1. November 1975
Nationalität	Pakistanisch

Studium

1994-1997	Studium der Biologie an der Universität zu Punjab, Pakistan
1998-1999	Diplomarbeit an der Quaid-i-Azam Universität, Islamabad, Pakistan, Abteilung "Biological Sciences "
1999-2001	M. Phil an der Quaid-i-Azam Universität, Islamabad, Pakistan Abteilung Biochemie/Molekulare Biologie, Thema: Genetic transformation of canola (*Brassica napus* L.) by *Agrobacterium tumefaciens*
2001-2005	Scientific officer am "Institute of Agricultual Biotechnology and Genetic Resources", NACR, PARC, Pakistan. Thema: Biotechnik der Pflanzen Raps, Olive, Reis, Banane, Kartoffel und Zuckerrohr
seit April 2005	Doktorarbeit am Max-Planck-Institut für Züchtungsforschung in Köln in der Abteilung Molekulare Pflanzengenetik (Prof. Dr. H. Saedler) Thema: Molecular evolution of the Inflated Calyx Syndrome (ICS) in Withania (Solanaceae)

Köln, den 4. Dezember 2008

Muhammad Ramzan Khan